Mapping Chromosomal Loci in Specific Language Impairment: A Pedigree-Focused Approach

By

Erin M. Andres

B.S., Regis University, 2013

Submitted to the graduate degree program in the Child Language Doctoral Program, and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Master of Arts.

_____

Chair: Mabel L. Rice, PhD

_____

Co-chair: M. Hashim Raza, PhD

_____

Erik A. Lundquist, PhD

Date Defended: 04 May 2018

The thesis committee for Erin M. Andres certifies that this is the approved version of the following thesis:

Mapping Chromosomal Loci in Specific Language Impairment: A Pedigree-Focused Approach

_____

Chair: Mabel L. Rice, PhD

_____

Co-chair: M. Hashim Raza, PhD

Date Approved: 11 May 2018

Abstract

Specific language impairment (SLI) is characterized by a delay in the mastery of language despite average or above average nonverbal intelligence (IQ). There are multiple assessments used in practice to measure the language abilities of individuals with SLI. Standardized language assessments in conjunction with a measure of nonverbal IQ are the most crucial for distinguishing individuals with and without SLI in research practice. Studies have found that the incidence of SLI in extended relatives of probands is significantly higher than population matched relatives of controls. The heritability estimates of SLI are higher in MZ twins than DZ twins. Both family and twin studies indicate genetic involvement in the transmission of SLI. Previous genetic studies in SLI have found candidate chromosomal loci on 2q24, 6p21, 10q26, 12p13, 21q, and several candidate genes including *TM4SF20, NFXL1, CNTNAP2, KIAA0319, CMIP*, and *ATP2C2* have been implicated in SLI. However, the causes of SLI are not well understood and investigation may benefit from family-based approaches. The current study approached genetic investigation of SLI one pedigree at a time. We report SLI loci on chromosomes 4q, 3p, 6q, 9q, 10q, 12p, 14q and 15q linked with the omnibus standard score categorical phenotype, indicating genetic and phenotypic heterogeneity of SLI. These findings support the discussion of previous hypotheses that SLI is a polygenic disorder, with multiple loci reported in a few of the families included in this report.

Keywords: genetic linkage, pedigree, language phenotypes, specific language impairment, reading phenotypes

Acknowledgements

I would first like to thank all of my committee members, Dr. Mabel L. Rice, Dr. M. Hashim Raza and Dr. Erik A Lundquist, for their contributions of time and feedback throughout this process. I would especially like to thank my advisor Dr. Raza, for his guidance in all lab protocols and pushing me to perform my best work. He was always able to trouble shoot issues with me in a way that made me feel like part of the process in finding the solution. I would also like to thank Dr. Rice for seeing potential and accepting me to the Child Language Doctoral Program, which has proved to be a wonderful place for both my passions and goals. She has been a wonderful guide through these early years as a woman in science. Thanks to the families who provided their DNA, and to Dr. Rice for the use of that DNA, which her lab has collected over the past 20+ years.

I would also like to acknowledge two of my undergraduate professors at Regis University. Dr. Mark Basham always encouraged me to pursue graduate school, even when I doubted myself, and Dr. Rebecca Betjemann led me to my current research topic through her seminar class “Learning Disabilities and Developmental Differences,” in which she emphasized the genotype-phenotype relationship.

Finally, I would like to thank my family, friends and fellow graduate students for their encouragement and always lending an ear to discuss both school-related and unrelated topics when I needed them.

Funding source: NIH grant - T32DC000052 – Training Researchers in Language Impairments

Table of Contents

List of Figures

List of Tables

List of Appendix A Tables

List of Appendix B Tables

Introduction

The U.S. National Institute of Deafness and Other Communicative Disorders (NIDCD, 2015) defines specific language impairment (SLI) as "a language disorder that delays the mastery of language skills in children who have no hearing loss or other developmental delays." An estimated 7% of the population has SLI, based on a prevalence study of kindergarten children in the U.S. (Tomblin et al., 1997). The majority of children acquire language as they develop without any formal instruction. Evidence shows that greater language ability correlates with multiple aspects of quality of life including stronger social relationships and academic success (Rice, 2017). Unfortunately, language does not come easily to all children. In the case of children with SLI, the cause of their language delay is not understood. It is important to note that individuals with SLI have average to above average non-verbal intelligence (IQ) (Rice, 2016; Rice, 2017).

Advances in the understanding of human genetics over the past few decades have allowed us to genetically investigate both simple Mendelian and non-Mendelian or complex traits (Jordan, 2014; van Heyningen & Yeyati, 2004). Classification of Mendelian traits is dichotomous, while complex traits are those that affect individuals on a continuous spectrum of phenotypes (Strachan, Goodship, & Chinnery, 2014). Mutations in a single gene can result in a disease phenotype and such mutations are transmitted from parents to offspring in a Mendelian pattern, (i.e. autosomal dominant, recessive or sex linked pattern) with complete disease penetrance (van Heyningen & Yeyati, 2004). Non-Mendelian traits are the result of mutations in single or multiple genes transmitted from parents to offspring under an undefined inheritance pattern with variable disease penetrance (van Heyningen & Yeyati, 2004). Non-Mendelian traits

are more complex to study genetically compared to Mendelian disorders, especially those that are phenotyped behaviorally (as SLI is) (Jordan, 2014).

Pedigrees provide a group of affected and unaffected related individuals who share phenotypes. The term phenotype, in genetics, refers to a set of observable or measurable inherited characteristics. Gene expression is known to control phenotypes, but there are hypotheses and evidence that environmental factors influence the genotype-phenotype relationship (Rutter, 2007). Theories of environmental influence are especially popular in complex genetic disorders, though evidence is not yet well established (Day & Sweatt, 2011; Rice, 2012). Despite individual differences across all measureable traits, evidence shows that variance of traits is reduced within families (Neale & Cardon, 2013). Individuals with well-characterized phenotypes can provide unique insight into the inheritance pattern within a family (Jordan, 2014). Pedigree-focused analysis (which is considered a form of genetic epidemiologic data) of complex Mendelian disorders with well-defined phenotypes can account for the variation in spectrum phenotypes (Neale & Cardon, 2013; Tomblin, Freese, & Records, 1992). Therefore, knowledge gained from a pedigree-focused approach with well-defined phenotypic measures can contribute to an understanding of the underlying molecular mechanisms involved in the transmission of complex traits.

Family history data provides evidence that the difficulties children with SLI experience are inherited (Rice, Haney, & Wexler, 1998; Tallal et al., 2001). Twin studies of SLI have revealed strong heritability estimates between 20-50% for dizygotic (DZ) twins and greater than 50% for monozygotic (MZ) twins (Bishop & Hayiou-Thomas, 2008; Rice, Zubrick, Taylor, Hoffman, & Gayán, 2018). Despite a wide range of heritability estimates across studies of SLI,

higher heritability estimates are consistent in MZ twins, indicating a strong genetic component in the transmission of SLI.

Complex disorders are less likely to fall into one of the Mendelian inheritance patterns (autosomal dominant or recessive, sex-linked dominant or recessive, co-dominant) (Strachan et al., 2014). Complex Mendelian disorders could show a combination of inheritance patterns within a family (van Heyningen & Yeyati, 2004). Family based parametric linkage analysis is considered a standard method to map disease loci in simple Mendelian disorders where mode of inheritance and disease penetrance is fully understood (Ott, Kamatani, & Lathrop, 2011). A lack of an established inheritance model and variable disease penetrance in complex Mendelian disorders influences parametric linkage analysis. Multigenerational large pedigrees with multiple affected and unaffected individuals may enhance the power of a pedigree for genetic investigation of a complex Mendelian disorder. In complex genetic disorders, this power is influenced by the variable disease penetrance, phenocopy rate and unknown segregation pattern (Clerget-Darpoux & Elston, 2007; Ott et al., 2011).

Both single gene and polygenic models have been proposed for speech and language disorders (Reader, Covill, Nudel, & Newbury, 2014). An example of a single gene theory comes from the identification of the *FOXP2* gene, located on chromosome 7q31, in the three-generational KE family of the United Kingdom (Fisher, Vargha-Khadem, Watkins, Monaco, & Pembrey, 1998). Originally, the teacher who identified the proband of the KE family thought the child had SLI (Fisher et al., 1998). In 1987, half of the members of the KE family were identified as having developmental verbal dyspraxia (Kang & Drayna, 2011). Once the phenotype of the KE family was defined more clearly as an oral motor difficulty, the idea of a single gene disorder was more plausible. In the case of SLI, many genes have been suggested, and multiple variants

have been identified within the same populations, providing support for polygenic theories. In polygenic disorders multiple genes interact together to cause a phenotype. One such example is the investigation of a large founder population on an isolated island near Chile, the Robinson Crusoe Island (RCI) founder population with over 100 individuals (Villanueva et al., 2011). Parametric linkage analysis has not resulted in a single co-segregating region for even the isolated Chilean population, providing support that SLI is polygenic (Reader et al., 2014). However, this could also indicate support for a conflicting or confounding idea that SLI is polyphenic.

Candidate genes identified from population cohorts include *CMIP* and *ATP2C2* on chromosome 16q (Newbury et al., 2009; Scerri et al., 2011) and *CNTNAP2* on chromosome 7q (a downstream target of the *FOXP2* gene) (Vernes et al., 2008; Whitehouse, Bishop, Ang, Pennell, & Fisher, 2011) and *TM4SF20* on chromosome 2q36 (Wiszniewski et al., 2013). The RCI founder population study identified the candidate *NFXL1* on chromosome 4p (Villanueva et al., 2015).

Multiple groups have reported linkage loci and candidate genes for SLI through linkage analysis, population, family based association, candidate gene approaches and next generation sequencing (Reader et al., 2014). Such studies provide further evidence for the involvement of genetic components in SLI (Paracchini, 2011; Rice, Smith, & Gayán, 2009). Though the literature of genetics of SLI reports many linkage loci, there is a lack of indisputable evidence attributing specific candidate loci or genes to a percentage of the population (Rice, 2012; Rice et al., 2009). This lack of consistency is in large part due to the complexity of SLI (Reader et al., 2014). Large cohorts of unrelated individuals with SLI, like the SLI Consortium cohorts in the UK, provided some of the first SLI linkage regions, specifically on chromosome 16q and 19q,

through the Haseman-Elston (HE)-multipoint and single point-regression analysis and the variance components (VC) method (SLI Consortium, 2002, 2004). Concurrent studies with Canadian and American families with SLI identified a third linkage region on chromosome 13q21 using parametric linkage analysis with both dominant and recessive models (Bartlett et al., 2004; Bartlett et al., 2002). However, this region linked with a slightly different SLI phenotype, which included participants with comorbid reading impairment (Bartlett et al., 2004; Bartlett et al., 2002). A few other studies have also reported linkage, using both parametric and nonparametric analyses, on chromosomes 2q24, 6p21, 10q26, 12p13, 21q (Reader et al., 2014).

The following genome-wide linkage analysis will focus on the behaviorally complex phenotype, specific language impairment (SLI). This study aims to show the advantages of pedigree-focused genetic investigation of a highly heterogeneous and poorly understood language phenotype-SLI. The current investigation intends to show how a genome-wide linkage analysis of extended families with well characterized SLI phenotypes and good quality SNP markers may help to map SLI genes in families ascertained as a part of a larger ongoing study at the University of Kansas (KU) in Mabel Rice's Language Acquisition Studies (LAS) lab.

Literature Review

Phenotype Characteristics of SLI

Currently, there is a lack of standardization of complex behavioral phenotypes, like SLI. A database is needed, like those developed for Mendelian diseases, i.e. Online Mendelian Inheritance in Man (OMIM). The NIH has funded the development of such a database, called PhenX (Phenotype Expression) (Rice & Tager-Flusberg, 2016). PhenX aims to provide a space for collection of diagnostic criteria for complex phenotypes. Presently, different diagnostic or phenotype measures for SLI are widely used in clinical and research practice. Even with the

development of PhenX, a large variety of assessments may continue to be used to measure SLI. Therefore, it is important to understand the assessments currently in use, in order to understand which trait measures are being targeted when a specific assessment is mentioned.

All types of genetic (association, linkage and twin studies) and behavioral investigation (familial aggregation, intervention, heritability) require a well-defined phenotype, no matter if it's a population-based or family-based study (Jordan, 2014; Rice et al., 1998; Tallal et al., 2001). A well-defined phenotype, in the case of a language or cognitive phenotype, would ideally utilize a limited number of the same assessments across studies and populations. Understanding the different phenotype assessments used to measure language abilities of children with SLI is crucial. For example, the UK group and their collaborators have produced many genetic results linked to the non-word repetition (NWR) task, which is known to measure phonological short-term memory (SLI Consortium, 2002, 2004). The practice of multiple assessments is common within the area of genetic analysis of phenotypes related to language and other cognitive abilities (Hanscombe et al., 2012; SLI Consortium, 2002, 2004).

Twin studies of behavioral phenotypes are extremely robust when it comes to the investigation of heritability (Neale & Cardon, 2013). MZ and DZ twin comparisons allow for a unique genetic investigation, and control of shared and unique environment (Neale & Cardon, 2013). However, the heritability estimates of twins with SLI continue to differ across studies (Bishop & Hayiou-Thomas, 2008; Rice, 2017). The range in heritability estimates reported can be attributed to many factors. Twin studies of SLI, from the same group over time, have attributed the differences in the estimates of concordance to the change in ascertainment of samples, criteria used for phenotyping and in the development of analysis methods (Bishop & Hayiou-Thomas, 2008). If the results of behavioral estimates vary across seemingly robust

genetic conditions, i.e. twins, controlling for the measures becomes even more important when using individuals who share less genetic material, or are even unrelated.

In order to ensure individuals meet the basic definition of SLI as defined by NIDCD (2015), articulation, verbal and non-verbal IQ must be assessed, for the purposes of behavioral and genetic studies. Generally, children with SLI do not show articulation deficits (Rice, 2018). Generally, their IQ is assessed using the age appropriate Wechsler Intelligence Scale, i.e. Preschool and Primary Scale (WIPPSI), the Scale for Children (WISC), or the Adult Intelligence Scale (WAIS). The majority of studies included in this review used the WISC and WAIS, with the exception of the Brisbane Adolescent Twin Sample (BATS), which used the Multidimensional Aptitude Battery to determine IQ (Luciano et al., 2013; Paracchini, 2011; Wright & Martin, 2004). The Token test can also assess general cognition. The Token test is designed to show an individual's ability to follow complex instructions. Participants from Canadian and American families with SLI were assessed with the Token test (Bartlett et al., 2004; Bartlett et al., 2002). An assessment like the Goldman Fristoe Test of Articulation (GFTA) measures articulation.

Generally, behavioral and genetic studies also include a standardized language test appropriate for the participant's age, i.e. an omnibus language score. Examples of the standardized assessments used (and the ages they are appropriate for) in the ascertainment of a cohort of samples from Kansas are listed. Children under 2;6 years were assessed with the Preschool Language Scale-3 – Total Language Score (PLS-3). Children between 2;6-3;11 years were assessed with the Test of Early Language Development-3rd Ed – Spoken Language Standard Score (TELD-3). Children between 4-6;11 years were assessed with the Test of Language Development-2 – Primary Spoken Language Standard Score (TOLD-2). Participants

older than 7 years were assessed with the Clinical Evaluation of Language Fundamentals-3rd Ed – Total Language Standard Score or Expressive Lang Score (CELF-3). Standardized language assessments, like those listed above, have many subtests in order to get a full picture of language ability. Examples of subtests on the CELF include sentence structure, phonological awareness, recalling sentences, and rapid automatic naming. It is more difficult to accurately measure grammar or syntax knowledge in adults. This makes it more difficult to measure adults' abilities in one of the hallmark deficits seen in individuals with SLI. There are parts of the CELF which measure grammar, but the majority of subtests are known to measure vocabulary and cognitive function associated with language.

The TELD-3, TOLD-2, and CELF-3 were used by the SLI Consortium, research groups in Kansas (Rice et al., 2009), Australia (Rice et al., 2018), and Canada (Bartlett et al., 2004; Bartlett et al., 2002), whose results are all included in this review. The Kansas group was the only one to use the PLS-3 (Rice et al., 2009).

Often reading assessments are included in studies of individuals with SLI. Common reading assessments include the Gray Oral Reading Test (GORT), and subtests of the Woodcock Reading Mastery Test (WRMT). The GORT assesses text reading and comprehension, and is generally used beginning at age 7 (Rice et al., 2009). The WRMT assess "word level reading" (Rice et al., 2009) or decoding skills. The WRMT has multiple subtests, Letter Identification is appropriate up to age 9, while Word Identification and Word Attack are used from reading age through adulthood (Rice et al., 2009). The WRMT was used in multiple studies presented in this review, while the Kansas group was the only study included here to use the GORT. The BATS cohort also included an assessment, which does not seem to be widely used, the Components of

Reading Examination (CORE). The CORE includes 120 regular-word, irregular-word and non-word reading and spelling items (Luciano et al., 2013).

A few studies have made unique decisions to distinguish individuals with and without SLI. For example, the studies completed concurrently with the SLI Consortium investigation (in the early 2000s) grouped their Canadian and American family participants into three phenotype categories: Clinical Impairment (CI-which accounts for adults who were never previously diagnosed and may have mastered compensatory skills), Language Impairment (LI) and Reading Impairment (RI) (Bartlett et al., 2004; Bartlett et al., 2002). It is important to note that no participants in the Canadian or American cohorts qualified exclusively as language impaired, though their intent was to investigate SLI (Bartlett et al., 2004; Bartlett et al., 2002). This was likely due to the inclusion of the CI category. In total, 25 participants were categorized as language impaired along with some combination of RI and CI. These participants completed the age appropriate Test of Language Development (TOLD), the Wechsler (WIPPSI, WISC, WAIS), and Token Test. They also completed the Word Identification and Word Attack subtests of the WRMT. Finally, self and parent report questionnaires were considered. The criteria for the LI group was a standard score less than or equal to 85 on the TOLD (Bartlett et al., 2004; Bartlett et al., 2002). RI was defined by a reading discrepancy score of a Word Attack standard score 1 SD below their Wechsler IQ score (Bartlett et al., 2004; Bartlett et al., 2002). The CI group was defined in multiple ways, but if the participants overall score was above 85 on the TOLD, but three of their subtest scores were below the mean of 7, or they scored less than 85 on the Token Test, they were defined as CI. Participants in the CI could also have average or above average scores, but have a history of language difficulty, specifically at least 2 years of speech language therapy (Bartlett et al., 2004; Bartlett et al., 2002).

The RCI founder population speaks Spanish. Often there are Spanish (unlike other non-English languages) versions of standardized assessments, like the Spanish version of the Peabody Picture Vocabulary Test (PPVT) (Villanueva et al., 2011). Notably, all inhabitants of RCI were first assessed with a non-verbal IQ measure, using the Columbia Mental Maturity Scale (Villanueva et al., 2011). Any inhabitants of RCI with hearing loss, oral motor difficulties or any comorbid diagnoses were excluded. Participants between the ages of 3 and 8 years, 11 months were assessed with an expressive and receptive morphosyntax measure (Toronto Spanish Grammar Exploratory test) and a phonology test (Villanueva et al., 2011). Older participants were classified on the basis of a family history interview, a verbal fluency test, verbal comprehension test and an auditory screening (Villanueva et al., 2011).

Vocabulary and grammar are two of the hallmark deficits seen in individuals with SLI (Rice et al., 2009; Rice et al., 2018). Longitudinal growth curves show evidence of delays in acquisition of vocabulary, aspects of morphosyntax and finiteness marking in individuals with SLI as compared to their age matched peers (Rice, 2012; Rice & Hoffman, 2015). Mean length of utterance (MLU) growth curves for children with SLI aligned with younger language-matched peers (Rice, Redmond, & Hoffman, 2006). Vocabulary, as measured using the Peabody Picture Vocabulary Test – 4th Edition (PPVT-4), is about two years behind age-matched typically developing (TD) peers (Rice, 2012).

Vocabulary. Individuals with SLI show deficits in receptive and expressive vocabulary (Rice, 2012). Receptive vocabulary refers to the vocabulary an individual comprehends while expressive vocabulary refers to language an individual can produce. Research has shown that IQ and vocabulary size correlate (Rice & Hoffman, 2015). Specifically, vocabulary assessment with the PPVT-4 is a robust tool that can be used effectively as a surrogate IQ test, or measurement of

underdeveloped vocabulary growth (Dunn & Dunn, 2007; Rice & Hoffman, 2015). Statistical comparison of this correlation to other measures of language acquisition, such as MLU, validates the correlation between IQ and vocabulary size (Rice et al., 2006). Children with SLI show delays in vocabulary acquisition as compared to their typically developing (TD) peers, when the PPVT is used as a latent trait (Rice, 2012). The delays not only remain through adolescence but the gap widens, as well (Rice, 2012). The evidence supporting the use of the PPVT-4 has made it a popular assessment in behavioral and genetic studies of SLI (Rice et al., 2009; Rice et al., 2018; Taylor, Christensen, Lawrence, Mitrou, & Zubrick, 2013; Wiszniewski et al., 2013).

Grammar. Individuals with SLI show a delay in morphosyntax, specifically inflectional morphemes (Oetting & Rice, 1993; Rice, Wexler, & Cleave, 1995). In TD children there is a period of language development called optional infinitive (OI) (Wexler, 1994). The OI stage is defined as a time when children do to not obligatorily mark tense or agreement for verbs (Wexler, 1994). An infinitive refers to the basic form of a verb; an example of not marking tense in English is a lack of '-ed' when referring to an action that occurred in the past. It was proposed that children with SLI show an extended OI (EOI) stage, meaning they exhibit this stage in their speech later in life than TD children. The EOI hypothesis aligns with the approximately two year delay shown by growth curves (Rice, 2012).

The Test of Early Grammatical Impairment (TEGI) was developed to more clearly distinguish children with and without language impairments (Rice & Wexler, 2001). The assessment is based on years of research conducted by Rice, including research within a language acquisition preschool made up of children with SLI, ESL, and TD children (Rice & Wexler, 2001). TEGI includes multiple subtests that make up the Elicited Grammar Composite: the Phonological Probe, main tests and Supplemental Probe. The phonological probe first

determines if the children are able to produce the sounds required. The other parts of the assessment are divided based on different types of inflectional morphemes and use of main verbs, i.e. third person singular, past tense and be/do. The supplemental probe uses grammaticality judgment, which looks at dropped markers and '–ing,' and agreement. Twin studies, from the Australian population, utilized multiple assessments (all mentioned in Participant section), and the TEGI showed the highest heritability estimates at 0.92, supporting the claim that grammatical delay is a hallmark of SLI (Rice et al., 2018).

Non-word repetition. The SLI Consortium, Avon Longitudinal Study of Parents and Children (ALSPAC) (Fraser et al., 2012; Newbury et al., 2009; Paracchini, 2011; Scerri et al., 2011; SLI Consortium, 2002) and BATS cohorts all completed a NWR task (Bishop, North, & Donlan, 1996; Luciano et al., 2013; Paracchini, 2011; Wright & Martin, 2004). The SLI Consortium used two versions of the Gathercole NWR task, one given to each set of participants (SLI Consortium, 2002, 2004). Downstream analysis determined that the two versions were correlated at 0.89 ($p < 0.001$) (SLI Consortium, 2002, 2004). The ASLPAC participants completed an adapted NWR task, with 12 items (evenly divided between 3, 4, and 5 syllable items) (Luciano et al., 2013). The BATS cohort used a standard score sum from two NWR tests: the Gathercole and Baddeley test and Dollaghan and Campbell test (Luciano et al., 2013). The group from Kansas also had their participants complete the NWR task as a part of the Comprehensive Test of Phonological Processing (CTOPP) (Rice et al., 2009). In the NWR task, a participant hears a non-word and repeats the non-word back. The non-words increase in the number of syllables as the test continues. The task is said to assess phonological short-term memory.

Genetic Studies of SLI

Family aggregation. As far back as 1880, researchers have investigated differences between familial and non-familial speech related disorders, such as stuttering (Stromswold, 2000). Familial aggregation refers to the clustering of physical or behavioral traits in related individuals. Familial aggregation studies provide the first set of evidence for the genetic transmission of traits. Modern familial aggregation studies use a case-control design, based on the ideas from the original investigations (Tallal et al., 2001).

A meta-analysis of 18 familial aggregation studies, completed between 1959 and 1996, showed variable rates of developmental language disorders in proband (17-78%) and control (3-16%) families (Stromswold, 1998). The meta-analysis lumped the phenotypes into a broad category of developmental language disorders because of the variable phenotypes across studies (Stromswold, 1998). Previous studies had variable phenotypes due to the changes in assessment and diagnostic criteria over time, as well as variable ascertainment procedures across studies (Stromswold, 1998).

Familial aggregation studies with more clearly defined SLI phenotypes have shown significant differences between the rate of SLI in families of probands with and without SLI (Rice et al., 1998; Stromswold, 1998, 2000; Tallal et al., 2001). Two studies in particular had more clearly defined SLI phenotypes.

In the first study, two biological parents and one biological sibling were available for all probands accepted into the study. All probands with and without SLI were ascertained at school, probands with SLI were receiving speech and language services at school (Tallal et al., 2001). The immediate family members (parents and siblings) of 22 probands with SLI (45 siblings, 44 parents) and 26 age-matched controls (33 siblings, 52 parents) all completed a family history

questionnaire and were behaviorally phenotyped (Tallal et al., 2001). The two groups did not significantly differ in their socioeconomic status (SES). Assessments included the age appropriate TOLD, Token Test and IQ test. According to affected status based on assessment, the rate of SLI in SLI proband families was 30%, while it was only 7% in control families, a significant difference. The rates were similarly significant when affected status was based on the family history questionnaire (32% in SLI proband families and 8% in control proband families) (Tallal et al., 2001).

The second study was made up of 31 probands showing clear grammatical deficits, specifically EOI, and 67 controls, as well as their immediate and extended family members. Probands were recruited between 4 years 6 months and 5 years 6 months of age, prior to entering kindergarten. Probands and controls were ascertained from preschools; again, probands were those individuals who were receiving speech language services. A portion of the probands came from a larger longitudinal study (Rice & Wexler, 1996). In total there were 555 family members of probands with SLI, and 1283 family members of controls included in the aggregation study (Rice et al., 1998). There were two groups of controls, language ability-matched controls (on average 2 years younger than the proband group) and age-matched controls (Rice et al., 1998). In order to assess the EOI stage in probands and controls, a combination of direct assessment and coding of spontaneous speech samples were evaluated (Rice et al., 1998; Rice & Wexler, 1996). The three groups significantly differed in accuracy of the EOI measures (Rice et al., 1998). Within the nuclear families there was a significant difference in the number of affected individuals; 22% of the family members of the probands with SLI were affected, while only 7% of the family members of the controls were affected. When the extended family members were included, the percentage of affected in the probands with SLI families went down to 15%, while

the control families only decreased to 6%; the difference between the groups was still significant (Rice et al., 1998).

Association studies. Association studies most often use a large group of unrelated individuals with and without a disease phenotype to investigate common and complex genetic traits (Ott et al., 2011). The popularity of genome-wide association studies (GWAS) (Ott, Wang, & Leal, 2015) and the availability of large cohorts have provided opportunities for association studies of cognitive and language traits (Fraser et al., 2012; Luciano et al., 2013). The data of large cohorts is especially useful for GWAS studies when birth records are available. The birth and health records, including a large amount of clinical data, available in Australia and the UK population, have been used for population-based association studies of SLI (Fraser et al., 2012). Association studies map risk alleles of small effect that can be attributed to the causation of a particular phenotype in the population. Information from large association analyses, with well-phenotyped individuals, can be a source of confirmation of previously reported chromosomal loci/associated alleles.

The Avon Longitudinal Study of Parents and Children (ALSPAC) is a general population cohort organized with the purpose of evaluating the impact of genes and environment on health and development (Fraser et al., 2012). All women who were expecting babies between April 1991 and December 1992, in a specified geographic region were able to participate in the study (Fraser et al., 2012). The mothers and their children continue to be assessed longitudinally. Using 211 families from the SLI Consortium and 490 individuals classified as having SLI from the ALSPAC cohort, about 1900 SNPs were screened that spanned 58 genes in the SLI locus on chromosome 16 (Newbury et al., 2009). The results indicated *CMIP* (chromosome 16q23.2-

q23.3) and *ATP2C2* (chromosome 16q24.1) were linked with phonological short-term memory measured by the adapted NWR task (ALSPAC) (Luciano et al., 2013).

Both the ALSPAC and BATS cohorts were analyzed with whole genome SNP genotyping. When the populations were combined in analysis, the 10 most significant SNPs associated with the NWR task were reported, despite each cohort completing different versions (Luciano et al., 2013). Two of the ten SNPs were on chromosome 16, along with three SNPs on chromosome 7 (Luciano et al., 2013). Four of the 10 SNPs were reported on chromosome 21, with three being located on the gene *ABCC13*. The allele frequency within both the ALSPAC and BATS cohorts for the SNPs on *ABCC13* was less than 0.07 (Luciano et al., 2013). *ABCC13* is a pseudogene located at chr21q11.2.

Pedigree based linkage studies in speech and language impairment. The purpose of parametric linkage analysis is to identify coinheritance of chromosomal loci with a phenotype (Read & Strachan, 2011). Linkage analysis is a powerful method when used in medium to large multigenerational families with multiple affected and unaffected individuals with well-characterized phenotypes (Strachan et al., 2014). In parametric linkage, factors such as model of inheritance, disease frequency, and disease penetrance must be set to complete the analysis. The mode of inheritance can be set to dominant or recessive with variable disease penetrance. Though the use of multiple models during analysis is not ideal, past studies have shown it produces powerful results for analysis of complex phenotypes (Abreu, Greenberg, & Hodge, 1999; Bartlett et al., 2002; Greenberg, Abreu, & Hodge, 1998). Despite the fact that inheritance of complex phenotypes is not expected to follow simple Mendelian patterns, past studies have shown that parametric analysis of complex phenotypes is more powerful in defining candidate loci, than nonparametric analysis.

Disease penetrance is the degree to which a gene (or a set of genes) is susceptible to causing a phenotype. Disease penetrance is also defined as the probability of an individual being affected with a specific phenotype given the individual's genotype (Ott et al., 2011). When everyone with the same phenotype has the same causative gene for a disorder, disease penetrance is 100%. Disease penetrance can be more clearly defined as the proportion of individuals who have a mutant allele(s) and express a set of symptoms by a predetermined age (Cooper, Krawczak, Polychronakos, Tyler-Smith, & Kehrer-Sawatzki, 2013). Highly penetrant disorders often follow Mendelian inheritance patterns. Whereas in genetic disorders of a variable disease penetrance, a mutant allele may not show susceptibility in affected individuals or vice versa in the same family.

The environment or other genes can influence the expression of disorders with low penetrance, which can make it hard to distinguish the effect of genes versus the environment on expression of the phenotype. The influence of environmental factors on language development can be studied under controlled conditions in twin studies. It has been shown that genes have a greater influence on language development than the environment (specifically, the influence of language input was studied) (Bishop, 2006; Rice, 2017; Rice et al., 2018). Single gene and polygenic models, as described in Table 1, have been proposed for SLI (Reader et al., 2014).

Table 1.

Mendelian versus non-Mendelian inheritance and their implications across methods.

	Simple Mendelian Inheritance	Complex Mendelian Inheritance
Phenotypic Classification	Dichotomous only -either affected or unaffected	Dichotomous or Continuous
Genetic Susceptibility	Single gene (most often)	Single gene or polygenic
Vertical Transmission	Mendelian transmission only -Autosomal dominant or recessive -Sex-linked dominant or recessive	Undefined inheritance pattern *still see aggregation in families and transmission from parents to offspring
Penetrance	Complete -All individuals with the disease variant express the same phenotype	Incomplete -Members of the same family may show variable disease penetrance
Pedigree-based methods	*Very powerful, even with sib pairs for the study of rare variants Traditional parametric linkage analysis Whole exome sequencing Whole genome sequencing	*Very powerful, with medium to large sized pedigrees for the study of rare variants Traditional parametric linkage analysis Whole exome sequencing Whole genome sequencing
Population – based studies	Uncommon	Common *Power is derived from the size of the cohort More successful with common phenotypes Genome-wide association studies Sib-pairs Twin studies

Pedigrees may also be analyzed using nonparametric linkage analysis, which does not require definition of mode of inheritance, disease penetrance or disease frequency. It could be the best choice for small pedigrees or sib pairs. The lack of parameters led researchers to believe it

was the best choice for complex phenotypes (Ott et al., 2011). Parametric linkage analysis provides more power compared to nonparametric linkage analysis when used in medium to large pedigrees, even when used with complex genetic disorders (Ott et al., 2011). The power of parametric analysis indicates it is the most appropriate choice for pedigree-based linkage analysis of complex phenotypes, as the alleles are likely inherited from recent ancestors.

Most often linkage analysis is performed genome-wide using hundreds and thousands of genetic markers with known chromosomal locations genotyped in available members of the family (Strachan et al., 2014). Linkage is statistically quantified in LOD scores. A LOD score is a logarithmic of odds representing the ratio of the probability of linkage for a defined recombination fraction (theta=0) (Strachan et al., 2014). A LOD score of 3.3 and above, at the recombination fraction zero, is considered evidence for significant linkage (p-value < 0.05) (Ott et al., 2015). A LOD score of -2.0 or below is considered evidence against linkage (Ott et al., 2015). The fundamental principle of linkage studies is that closer alleles are more often co-inherited together, as a haplotype, and therefore it is less likely that these alleles are separated by recombination in multiple meiosis (Strachan et al., 2014). Chromosomal segments with genetic markers that show the highest LOD scores can be used to identify candidate genes. Two-point linkage analysis refers to the algorithm that determines if two markers (genetic location and a disease phenotype) segregate together in a family (Ott et al., 2015).

The SLI Consortium, in the UK, completed the first linkage analysis in individuals with SLI (SLI Consortium, 2002). The SLI Consortium sample was originally made up of 98 families, parents and siblings of probands contributed DNA and were evaluated behaviorally (SLI Consortium, 2002). The same group followed up with an additional 86 families collected from multiple hospitals around the UK (SLI Consortium, 2004). The UK group presented the linkage

results of a quantitative-trait locus (QTL). Two regression-based linkage analysis methods were used to analyze the genotypes: Haseman-Elston (HE) analysis and variance-components (VC) analysis. These linkage analyses rely on sib-pair information; therefore, no parental phenotype information was included in either analysis. The resulting loci were on chromosome 16q24, linked with the NWR task, and chromosome 19q13, linked with the CELF-R expressive language task (SLI Consortium, 2002). The resulting LOD scores were ≥ 2.2 for both regions for four analyses types: single-point and multipoint HE and VC analyses (SLI Consortium, 2002).

The follow-up study was targeted the resulting loci (chromosome 16q and 19q) (SLI Consortium, 2004). The linkage regions were replicated in the 86 new families using 40 polymorphic microsatellite markers on chromosome 16, and 19 (SLI Consortium, 2004). However, in the follow-up study both linkage regions (chromosome 16q and 19q) were linked to the NWR task, there was no linkage with CELF-R (SLI Consortium, 2004).

Five large Canadian families (Bartlett et al., 2002), and an additional 22 American families (Bartlett et al., 2004) were divided into the RI, CI, and CI+LI groups (as detailed in the 'Phenotype Characteristics of SLI' section). The five large Canadian families were originally ascertained through a Schizophrenia study and analyzed in branches. There were 86 total Canadian individuals, 73 of which were phenotyped and divided into the RI, CI and CI+LI groups (Bartlett et al., 2002). A follow-up study by the same group included 22 additional American families. Participants were genotyped with 381 microsatellite markers spanning the whole genome at 9 cM spacing, with an average heterozygosity score of 0.76, from the Weber Screening Set v 6.0 (Bartlett et al., 2002).

The most significant linkage was revealed at 13q21 with a LOD score of 3.92 for individuals identified with RI (Bartlett et al., 2002). The researchers hypothesized that the RI

group showed a reading deficit as a result of an underlying language impairment (Bartlett et al., 2002). Their hypothesis came from the fact that chromosome 13 had never been identified in individuals with reading disorder (RD). However, this study was completed in the early stages of genetic investigation of SLI, prior to the reports of a large number of studies showing varied results across populations, even with groups that were similarly phenotyped. There was also suggestive linkage at 2p22 and 17q23 for those in the CI + LI group (Bartlett et al., 2002). A follow-up investigation included microsatellite genotyping on chromosomes 2, 7, and 13 in the previous Canadian samples and American samples (Bartlett et al., 2004). Chromosome 7 was included to investigate convergence with an identified Autism Spectrum Disorders (ASD) locus. When the 22 American families were included in analysis, there was a LOD score > 4.0 at two locations on chromosome 13q (Bartlett et al., 2004). However, the inclusion of the American families did not confirm any loci on chromosomes 2 or 7 (Bartlett et al., 2004).

Founder populations and consanguineous families provide another unique opportunity for family-based analysis. One such founder population with a high incidence of SLI has been investigated: the Robinson Crusoe Island (RCI) (near Chile) founder population (Villanueva et al., 2011). The RCI population originated from a few founders. As of 2008, 77% of the inhabitants were related to the founders. Both parametric and nonparametric linkage analyses, along with homozygosity mapping were completed in 111 related individuals (44 of which had SLI) (Villanueva et al., 2011). Parametric linkage under dominant and recessive models of inheritance with the disease frequency set to 0.35 (determined by the RCI population) did not warrant any suggestive or significant linkage loci (Villanueva et al., 2011). Nonparametric linkage analysis resulted in five significant linkage loci, including: chromosome 6q, 7q, 12, 13 and 17 (Villanueva et al., 2011). Family based genome wide association study, completed with

the same sample, revealed a significant maternal parent-of-origin effect on chromosome 5p13 and suggestive paternal parent-of-origin effect on chromosome 14q12 (Nudel et al., 2014). Follow-up whole exome sequencing (WES) in > 100 members of this family revealed coding SNPs associated with SLI, on the *NFXL1* gene, located on chromosome 4p12 (Villanueva et al., 2015).

A large Cameroonian pedigree with five large branches was investigated using methods similar to the current study (Raza et al., 2013). The Cameroonian pedigree had a high prevalence of persistent stuttering (71 total family members-36 affected, 6 unknown). Parametric linkage using microsatellite and SNP genotyping was used to find loci co-segregating with persistent stuttering. Multiple linkage analyses were performed in this family and multiple stuttering loci resulted in one family, suggesting polygenic inheritance (Raza et al., 2013). The narrowed down linkage regions allowed for further analysis using whole exome sequencing and reported a rare heterozygous variation on the gene AP4E1 that co-segregated in one part of the family and in population-matched stuttering cases (Raza et al., 2015).

Previous Reports in Participants of Interest

A previous study of individuals with SLI, from the ongoing longitudinal study in the LAS Lab at KU, used reported loci for targeted linkage and association analysis (Rice et al., 2009). Chromosomal loci on 1p36, 3p12-q13, 6p22, and 15q21 reported in RD were investigated in the Kansas SLI cohort. A chromosomal locus, 7q31, which was reported previously as a candidate for speech and language impairment (*FOXP2* locus), was also included in the investigation. Many individuals from the families included in current study were included in this investigation of targeted loci. However, in the previous study the family relationships were not included as a parameter. In total 322 individuals were included in the investigation; 86 probands, 134 sibs and

102 parents and other relatives, meaning there were not two parents for every proband (Rice et al., 2009). Microsatellites across the five previously reported chromosomal regions were used for linkage analysis, and association was completed with SNPs for two of the previously reported regions (chromosomes 6p22 and 7q31) (Rice et al., 2009).

Linkage analysis was performed using quantitative and categorical phenotypes. For the quantitative phenotypes, the MERLIN-regress model was used to perform linkage analysis. Non-parametric linkage analysis was performed in MERLIN for the categorical phenotypes. LOD scores above 0.6 were found for the GORT categorical, and omnibus categorical and quantitative phenotypes within the chromosome 1p36 locus (Rice et al., 2009). LOD scores above 1.0 were found for the categorical PPVT phenotype on chr3p12-q13, which contains the RD candidate gene *ROBO1* (Rice et al., 2009). LOD scores above 0.6 were found for the TEGI quantitative and categorical, omnibus quantitative and categorical, GORT categorical, CTOPP categorical and MLU quantitative phenotypes on chromosome 6p22, which contains the RD candidate genes *DCDC2, KIAA0319* and *DYX1C1* (Rice et al., 2009). LOD scores above 0.6 were found for the omnibus quantitative language score phenotype on chromosome 7q31 (Rice et al., 2009). LOD scores above 0.6 were found for almost all of the phenotypes (categorical and quantitative WRMT, omnibus, CTOPP phenotypes, and the quantitative GORT and GFTA phenotypes) on chromosome 15q21 (Rice et al., 2009).

Association analysis was performed with quantitative phenotypes using quantitative transmission-disequilibrium test (QTDT) and family-based allelic association tests (FBAT), across 36 selected SNPs on chromosome 6p22 and 17 SNPs on chromosome 7q31 (Rice et al., 2009). Using QTDT, there was a significant association with GFTA (4 SNPs) and GORT (2 SNPs) on chromosome 6p22. Under FBAT analysis there was a significant association with

GFTA (2 SNPs), PPVT (1 SNP), GORT (3 SNPs), and omnibus (2 SNPs) (Rice et al., 2009). After QTDT analysis, there was a significant association with one SNP outside of the *FOXP2* gene on chromosome 7 with WRMT, GORT and omnibus. With FBAT analysis there was a significant association on chromosome 7q31 with GFTA (3 total SNPs, 1 on *FOXP2*), and the omnibus language score (2 total SNPs, 1 on *FOXP2*) (Rice et al., 2009).

Summary and Broader Lessons of the Previous Literature

Previous studies have shown limited convergence in the reported regions linked and associated with SLI (Table 2). When whole genome investigations yield results for a specific phenotype as they did for the SLI Consortium cohort (chr16q and 19q), follow-up targeted linkage of the resulting regions, with individuals assessed in the same way, results in replicated loci (Table 2) (SLI Consortium, 2002). The same was true for the investigation of Canadian families categorized as RI, LI and CI, who showed linkage with a locus on chromosome 13q; subsequent follow-up investigation of American families phenotyped by the same researchers replicated the significant locus (Bartlett et al., 2004; Bartlett et al., 2002).

Previous studies emphasized the NWR task phenotype, following the reports by the SLI Consortium. The NWR task is easily administered, therefore when general population cohorts were collected in Australia (ALSPAC and BATS), the Gathercole NWR task and adapted versions were added to the protocol. Studies utilizing these cohorts showed association with NWR on chromosome 16 and 7, which were previously reported, as well as novel genes, specifically on chromosome 21 (*ABCC13*) (Luciano et al., 2013).

The investigation of the RCI founder population showed the benefits of a pedigree-focused approach (Villanueva et al., 2011). The lack of a single causative variant emerging from the RCI population could indicate SLI is polygenic and/or polyphenic.

Table 2.

Summary of candidate loci and genes from previous SLI genetic studies.

Finding	Method	Phenotype	Population	Reference
CMIP, ATP2C2 (chr16)	Association (targeted)	NWR (adapted) - 12 items (3 each – 3, 4, 5 syllables)	ALSPAC (mother-child general population cohort)	(Fraser et al., 2012; Luciano et al., 2013; Newbury et al., 2009)
ABCC13 (3 variants)	Association (SNPs)	NWR (sum std score: Gathercole & Baddeley + Dollaghan & Campbell)	BATS (Australian twin cohort)	(Luciano et al., 2013)
Chr19q13, Chr16q24 + Chr16q, 19q	Linkage + targeted follow-up	CELF NWR (2 versions of Gathercole)	Sib pairs	(SLI Consortium, 2002, 2004)
Chr13q, 2p22, 17q23 Chr13q	Linkage + follow-up targeted	RI – reading, LI – language, CI – clinically impaired	Canadian and American nuclear families	(Bartlett et al., 2004; Bartlett et al., 2002)
Chr6q, 7q, 12, 13, 17 rare variant on *NFXL1* (4p12)	NPL WES	Standardized grammar + vocab assessments (in Spanish) +Non-verbal IQ	RCI founder population	(Villanueva et al., 2011; Villanueva et al., 2015)
Chr1p36, 3p12-q13, 6p22 & 15q21	Targeted linkage and association	Comprehensive (Non-verbal IQ, vocab, NWR, reading, etc.)	Kansas cohort	(Rice et al., 2009)

NPL = nonparametric linkage
WES = whole exome sequencing

Research Objectives

The principle aim of this investigation is to test the logic of pedigree-focused genetic analysis of SLI. The first research objective is to map chromosomal loci through genome-wide parametric linkage analysis using SNP markers on the Illumina Infinum QC Array-24 in six families with SLI ascertained from Kansas. The second objective is to perform microsatellite genotyping of the resulting linkage loci and perform further linkage analysis. The results will be useful for the investigation of genetic variants responsible for SLI, through the analysis of next-generation sequencing data in these families.

Method

Participants

The SLI families included six probands, 12 parents, 28 siblings, and 14 other relatives (family 2 is the largest family with 12 extended relatives). Affectedness status in this study was assigned based on the omnibus standard score phenotype. All six probands had a standard score one standard deviation below the average, as well as 32 other family members, totaling 38 affected individuals. All six families are Caucasian and are part of a larger ongoing longitudinal study in Dr. Mabel Rice's LAS Lab at KU. Figure 2a-f shows pedigrees of the six SLI families included in this study. For SNP genotyping, there were a total of 7 individuals available in Family 1, 18 in Family 2, 7 in Family 3, 10 in Family 4, 7 in Family 5 and 11 in Family 6.

The majority of the SLI families were recruited from school speech pathology caseloads, though family 2 self-referred to the study. The families were then assessed to ensure they met the requirements of the study and probands were carefully screened to ensure they met exclusionary criteria. The institutional review board (IRB) at the University of Kansas approved the study and

we acquired appropriate informed consent from all participants. Parents provided consent for all participants under 18 years of age.

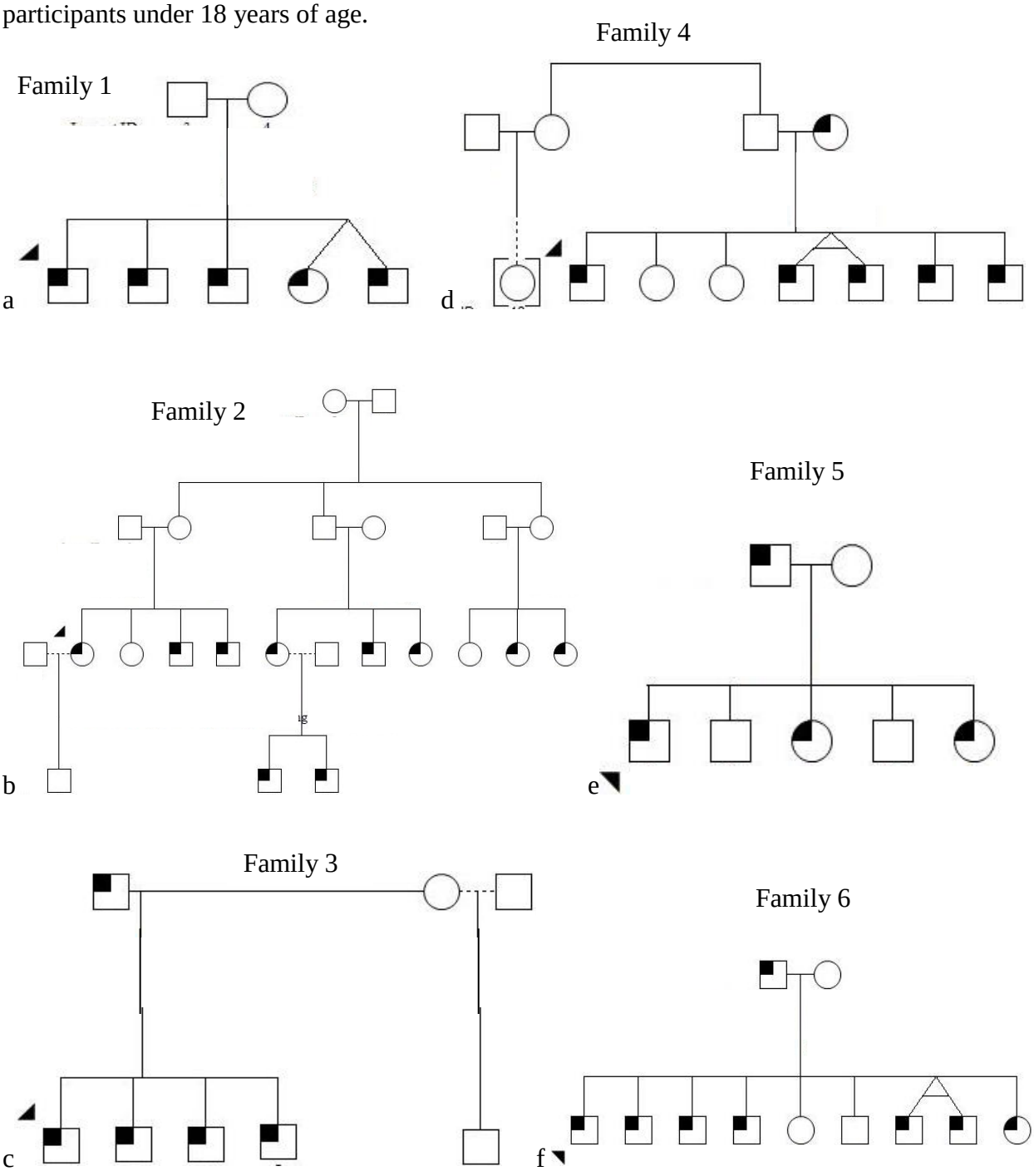


Figure 1. Pedigrees of families with SLI included in this study.
Family 1 (a). Family 2 (b). Family 3 (c). Family 4 (d). Family 5 (e). Family 6 (f).

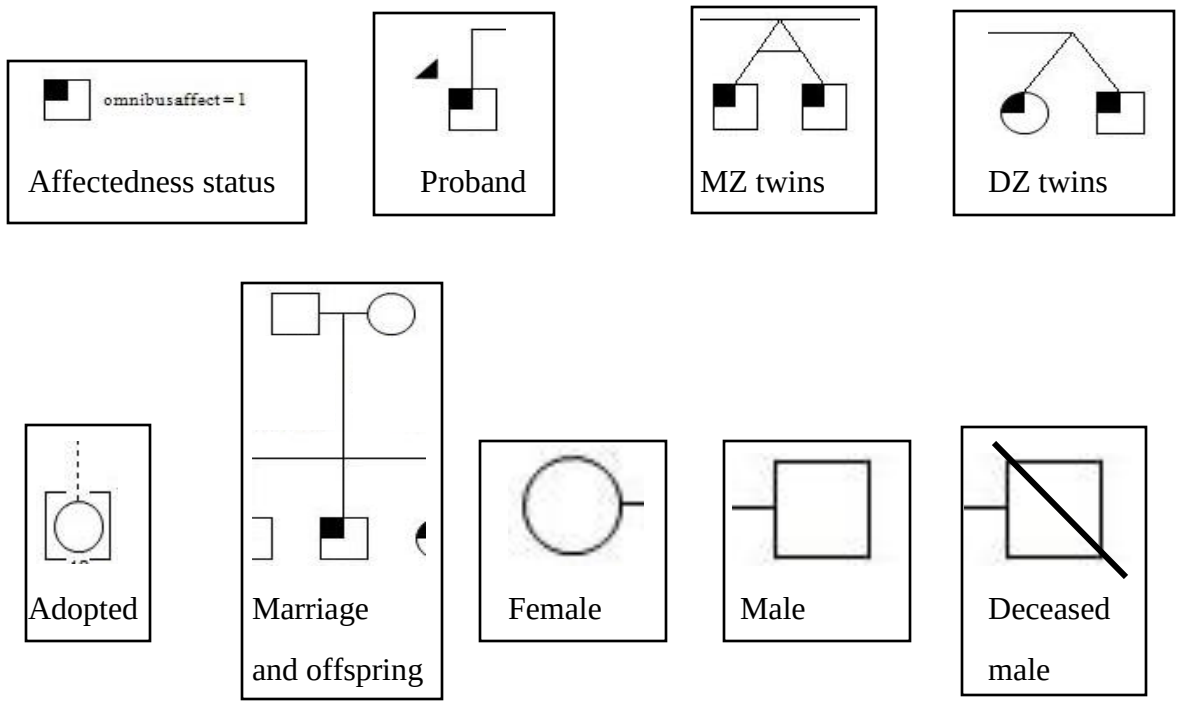


Figure 2. Pedigree key.

Measurement of language phenotype. Phenotypic measurements of participants from the longitudinal study of Kansas families are very clearly described by Rice and colleagues (2009). Probands in the Kansas families have no hearing loss, no cognitive impairment (average nonverbal intelligence), and no behavioral diagnoses at the initial time of assessment (based on parent report or a review of medical records). Probands were required to be monolingual speakers of standard American English dialect with clearly impaired language abilities. Speech articulation was assessed with the GFTA standard score.

The omnibus scores were obtained from a standardized language test appropriate for the participant's age. Within this sample, the PLS-3, TELD-3, TOLD-2 and CELF-3 were used. Affectedness status was assigned based on the standard score. If the child was one standard deviation below the average score, they were assigned as affected for that phenotype. In the following linkage analyses, the omnibus language score phenotype was used to determine

affected individuals. Individuals with a standard score ≤ 85, on the age-appropriate omnibus measure, were assigned affected for the omnibus language score phenotype. The participants under 9 years old were assessed with seven additional measures, while older participants completed five additional measures. The affected status for these measures was not utilized in these analyses. The additional measures are clearly described by Rice and colleagues (2009).

Genetic Analyses

Saliva samples from all participants were collected using the Oragene-Discover OGR-500 Kit from DNA Genotek (Oragene) at the time of behavioral assessment. DNA was purified from the saliva samples at University of Nebraska, using a standardized protocol with the prepIT kit provided by Oragene. Extracted DNA was re-suspended and stored in 1x Tris-EDTA (1X TE) buffer at -80˚C for long-term storage.

SNP genotyping. SNP genotyping was performed on the Illumina Infinum QC Array-24 that has 15,949 SNP markers equally spaced throughout the genome (https://www.illumina.com/products/by-type/microarray-kits/infinium-qc.html). All DNA samples were diluted with 1X TE buffer to 50ng/ul concentration and 200 ng DNA was used for SNP genotyping. SNP genotyping was outsourced to Johns Hopkins University School of Medicine, Genetic Resources Core Facility.

Linkage analysis. Whole genome parametric linkage analysis was used to identify candidate chromosomal loci in six SLI families. Linkage was performed using Superlink-Online SNP 1.1, a publically available genetic linkage analysis program at http://cbl-hapw.cs.technion.ac.il/superlink-snp/ (Silberstein et al., 2012). Superlink-Online SNP 1.1 has a graphical interphase, allowing it to handle large pedigrees and thousands of SNP markers.

The program requires two files: a data file (SNP file) and a traditional pedigree file in linkage format (ped file). The SNP file contains information about the markers (i.e. rsIDs, chromosomes and base pair positions), along with genotype columns for each family member. The ped file is a pedigree file in a linkage format that provides family information with relationships of each individual, their sex and affection status. Superlink-Online SNP 1.1 was used to perform two-point parametric linkage analysis in the Kansas SLI families (Silberstein et al., 2012). Linkage analysis was performed in all the families using both Mendelian models of inheritance (dominant and recessive), which was simply set within SuperLink. For each model, an analysis was run at two disease penetrance levels, complete (0, 0.99, 0.99 (dominant) and 0, 0, 0.99 (recessive)) and reduced (0, 0.70, 0.70 (dominant) and 0, 0, 0.70 (recessive)). A rare disease allele frequency of 0.001 was used in all analyses.

Following linkage analysis of the whole genome, regions of interest were selected for follow-up microsatellite genotyping. Regions of interest were defined based on the highest LOD scores observed in the whole genome. The boundaries for the regions of interest were 5 cM on either side of the highest LOD score in the region (or if all of the SNP markers with high LOD scores shows the same maximum LOD score, the boundary extended 5 cM from the marker closest to the telomere and centromere). The Marshfield Map was used to select microsatellite markers. Eighteen microsatellite markers across eight regions of interest were selected using the UCSC genome browser hg19 assembly. Microsatellite markers are short tandem repeat polymorphisms, developed based on extensive screening of recombination (Broman, Murray, Sheffield, White, & Weber, 1998). The markers have standardized primers for amplification found in the UCSC Genome Browser (Broman et al., 1998; Rhead et al., 2009). Oligo sequences of all markers synthesized through Eurofins Scientific. All forward primers were fluorescently

labeled on the 5' end with 6-FAM dye. The heterozygosity scores and physical positions of the microsatellite markers are presented in Table 4. PCR amplifications were performed in 10 ul reaction volumes (Table B1) using thermocycling programs (Table B2) and the reaction conditions are presented in the appendix. The appendix also includes the protocol used to prepare the reactions for capillary electrophoresis in the AB3130 xl genetic analyzer. The Applied Biosystems (ABI) GeneScan 500 LIZ Size Standard was used as the size standard during capillary electrophoresis. GeneMapper5 software (https://www.thermofisher.com/order/catalog/product/4475073) from ABI was used to analyze the genotyping data generated by the 3130xl Genetic Analyzer.

Multipoint Engine for Rapid Likelihood Inference (MERLIN) (http://csg.sph.umich.edu/abecasis/merlin/tour/parametric.html) is another computer program for linkage analysis, freely available online. MERLIN also requires multiple files, similar to SuperLink and was used to run parametric single point and multipoint linkage analysis as well as non-parametric linkage analysis (Abecasis, Cherny, Cookson, & Cardon, 2002). Specifically, it required a pedigree file, a data file, a map file and a model file (Abecasis et al., 2002). The map file contained the information about markers and their genetic or physical location. The pedigree file for MERLIN included the family relationship information and the genotypes for all markers. MERIN is limited in its ability to analyze large pedigrees but not in the number of markers that it can analyze.

Single point parametric linkage analysis of targeted loci obtained through Superlink was completed in MERLIN following genotyping of microsatellite markers in SLI families, with the goal of obtaining the single point LOD scores for the microsatellite markers in the linkage region. The analysis parameters used in Superlink were consistent with those applied in

MERLIN. MERLIN was used to produce multipoint and non-parametric LOD scores for the targeted loci.

Informative SNP markers in the linkage regions identified through SuperLink, were analyzed under both modes of inheritance in MERLIN. The resulting uninformative markers could be different under different parameter. For family 1, the genotypes for the targeted locus at chromosome 4q were extracted from the SNPs in the recessive SuperLink output. In family 2, genotypes for all three loci (chromosomes 3p, 9q and 15q) were extracted from the SNPs present in the dominant SuperLink output. For family 4 genotypes for all three loci (chromosomes 6q, 10q and 12p) were extracted from the SNPs present in the recessive SuperLink output. In family 6, all of the genotypes for the targeted locus on chromosome 14q were extracted from the SNPs in the recessive SuperLink output.

Results

Verification of Inheritance within Pedigrees

Quality control. The Genetic Resources Core Facility at Johns Hopkins University School of Medicine released genotype data to our lab in a file format suitable for GenomeStudio 2.0.3 by Illumina. Genotype data was consistent in all individuals with a call rate above 99.8%.

Single Point Parametric Linkage Analysis

The LOD scores for all SNP markers in the targeted loci, calculated by MERLIN, for both the recessive and dominant single point analyses are presented in Appendix A. Family 1 output is presented in Table A1 (chromosome 4). Output for family 2 is shown in Table A2 (chromosome 3), A3 (chromosome 9) and A4 (chromosome 15). Family 3 output is found in Table A5 (chromosome 6), A6 (chromosome 10) and A7 (chromosome 12). Finally, family 4 output are shown in Table A8 (chromosome 14).

Family 1. Single point parametric linkage analysis of family 1 in SuperLink showed evidence of suggestive linkage on chromosome 4q under the recessive inheritance model with complete penetrance. The 36.7 Mb chromosomal region identified on 4q31.23-q35.2 (rs1394845-rs6553000). A maximum LOD score of 2.4 was found on a SNP marker, rs6536541 (4q32.1) (Table 3). Several markers in the neighborhood of rs6536541 showed a LOD score of 1.2 (Figure 3a). When the model of inheritance was set to dominant, only one marker showed a LOD score above 1.0. It was the same maker with the highest LOD score under the recessive model; rs6536541 showed a LOD score of 1.1976 (Figure 3b and Table A1). Graphs produced by MERLIN in the targeted loci are shown in Figure 3c (recessive) and 3d (dominant).

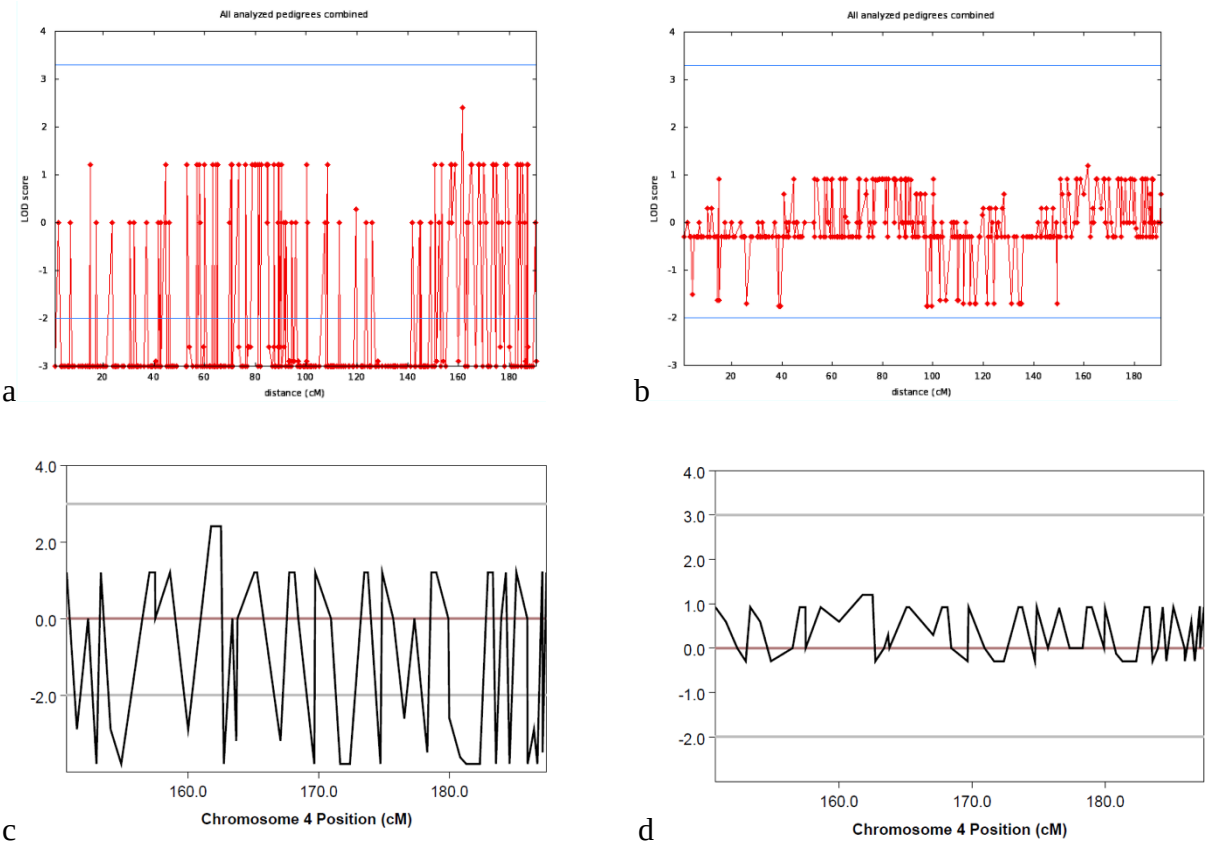
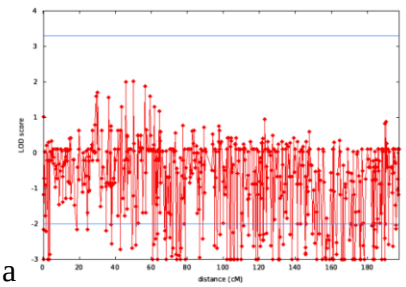
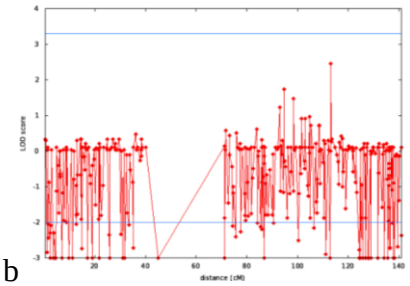
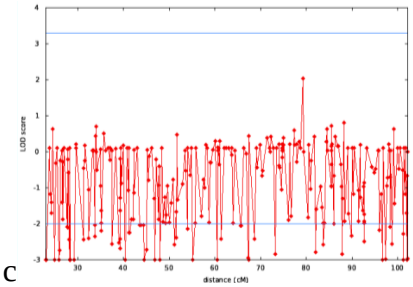
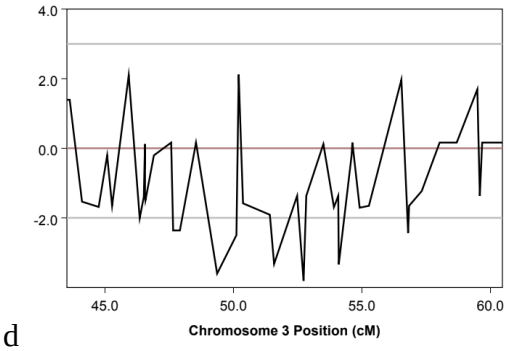


Figure 3. Chromosome 4q locus in family 1.
Parameters: omnibus affected, full penetrance, disease frequency: 0.001. Superlink graphs with recessive inheritance (a) and dominant mode of inheritance (b). MERLIN parametric linkage output, for chromosome 4q31.23-35.2 (150.7 – 187.42 Mb) with recessive inheritance (c) and dominant mode of inheritance (d).

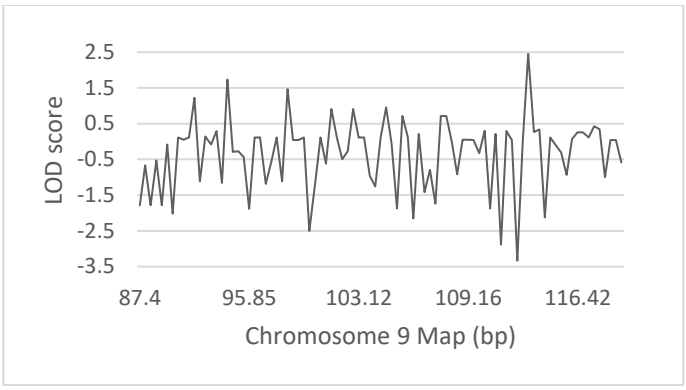
Family 2. Single point parametric linkage analysis of family 2 in SuperLink showed evidence of suggestive linkage on chromosome 3p22.1-p14.2 (Figure 4a), chromosome 9q21.33-q33.1 (Figure 4b) and chromosome 15q24.1q25.2 (Figure 4c) under the dominant model with complete disease penetrance.

A LOD score above 2.0 was obtained on two SNP markers in 16.9 Mb region at chromosome 3p22.1-p14.2 (rs1351631-rs1447971) in family 2 (Figure 4a). The majority of surrounding markers show LOD score over 1.0 under the dominant model with complete disease penetrance. A maximum LOD score of 2.4519 was obtained at chromosome 9q31.3 (rs984071) (Table 3). SNP markers rs3181360- rs1923433 defined the 30.6 Mb region on chromosome 9q21.33-q33.1 (Figure 4b). A maximum LOD score of 2.0267 was obtained on rs11072823 (chromosome 15q25.1) (Table 3) within the suggestive locus of chromosome 15q24.1-25.2 and multiple markers showed a LOD score above 1 (Figure 4c). However, the same marker on chromosome 15 (rs11072823) was found to have a LOD score greater than 3.0 when the family 2 analyzed under the recessive inheritance model with complete disease penetrance (Table 3). Graphical representation of the LOD scores in MERLIN shown in Figure 4d (3p22.1-p14.2), 4e (9q21.33-q33.1) and 4f (15q24.1q25.2).

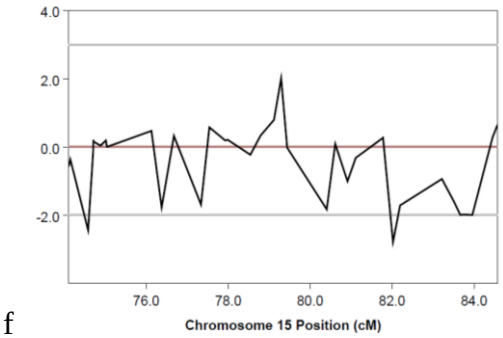
a  b  c 



d



e



f

Figure 4. Linkage results for loci in Family 2.
Parameters: omnibus affected, dominant inheritance, full penetrance, disease frequency: 0.001. Superlink graphs at chromosome 3 (a) and chromosome 9 (b) chromosome 15 (c). MERLIN parametric linkage graphs at chromosome 3p22.1-14.2 (43.5-60.5 Mb) (d) and at chromosome 9q21.33-33.1 (87.39-118.0 Mb) (e) at chromosome 15q21.1-25.2 (74.1-87.57 Mb) (f).

Family 3. Single point parametric linkage analysis in SuperLink did not show evidence of linkage under dominant or recessive models in family 3. All LOD scores were less than 1.0 (data not shown).

Family 4. Single point parametric linkage analysis in SuperLink showed evidence of suggestive linkage on three chromosomes 6q, 10q and 12p under the recessive mode of inheritance with complete disease penetrance. A LOD score of more than 2.5 (Table 3) was obtained on all three chromosomes under the recessive model (Figure 5a-5c).

A maximum LOD score of 2.6522 was obtained at rs9374570 and a chromosomal region of 13.08 Mb was defined as 6q21-q22.3. Several markers in the neighborhood show a LOD score above 1.4. (Figure 5a). Maximum LOD scores of 2.65 was found on multiple markers at

chromosome 10q21.1-q25.1 (51.2 Mb). There were several other markers in this region showed LOD scores above 1.0. There were nine markers in the entire 51.2 Mb region that showed negative LOD scores (Figure 5b). A maximum LOD score of 2.65 was identified in the 19.8 Mb region on chromosome 12p13.31-p11.2 on rs1366 (chromosome 12p12.1) (Figure 5c). Several markers in the neighborhood showed a LOD score above 1.4, while only two markers showed a negative LOD score. The LOD scores reduced to less than 2.0 on chromosome 6q, 10q and 12p when linkage analysis was performed under the dominant inheritance model (data not shown). Graphs produced by MERLIN for the only the SNP markers in the targeted loci are shown in Figure 5d (6q21-q22.3), 5e (10q21.1-q25.1) and 5f (12p13.31-p11.2).

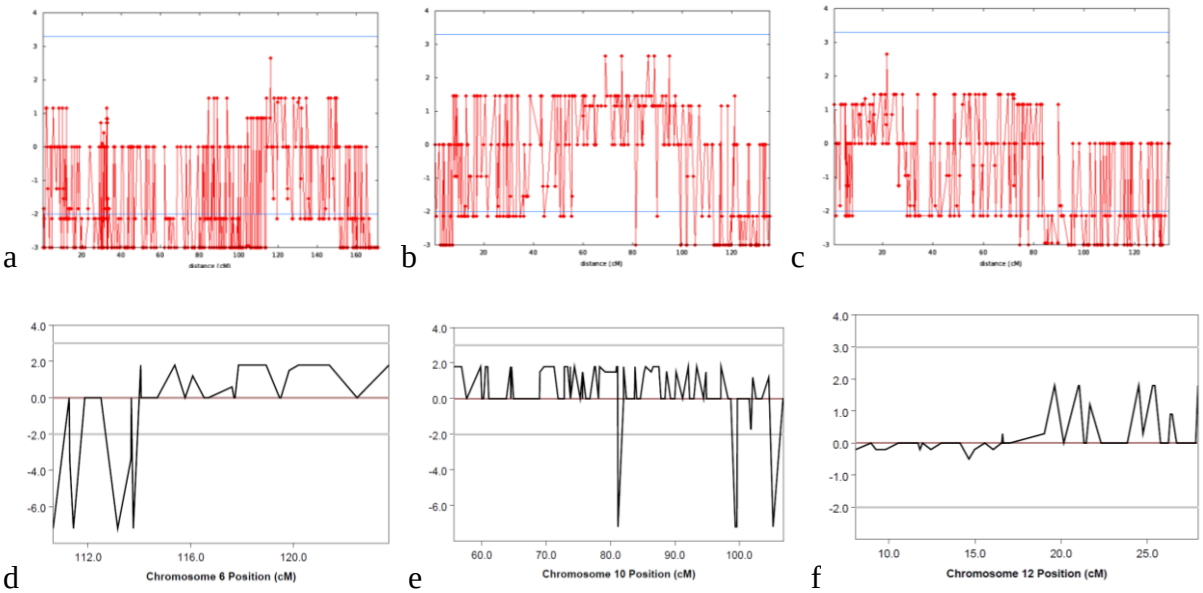


Figure 5. Linkage loci in Family 4.
Parameters: omnibus affected, full penetrance, disease frequency: 0.001. Superlink graphs for recessive mode of inheritance at chromosome 6 (a), chromosome 10 (b) and chromosome 12 (c). MERLIN parametric linkage graphs at recessive inheritance chromosome 6q21-q22.3 (110.6-123.7 Mb) (d) and chromosome 10q21.1-q25.1 (55.6-106.9 Mb) (e) chromosome 12p13.31-p11.2 (8.5-27.3 Mb) (f).

Family 5. No linkage obtained in family 5 when parametric linkage analysis was done under dominant and recessive mode of inheritance.

Family 6. Single point parametric linkage analysis identified evidence of suggestive linkage under the dominant inheritance model on chromosome 14q11.2-q21.1 in family 6. Several markers with LOD scores of 1.5 were found on chromosome 14 (Figure 6a and Table A8). A maximum LOD score of 2.0 was identified on several SNPs ranging from rs11628338 (chromosome 14q11.2) to rs368181 (chromosome 14q13.2) (Table 3). The MERLIN graph under the recessive model for family 6 (Figure 6d) included additional SNP markers toward the telomere end of 14q compared to the graph for the dominant model (Figure 4c). The additional markers were excluded from Table A8 because they all showed negative LOD scores.

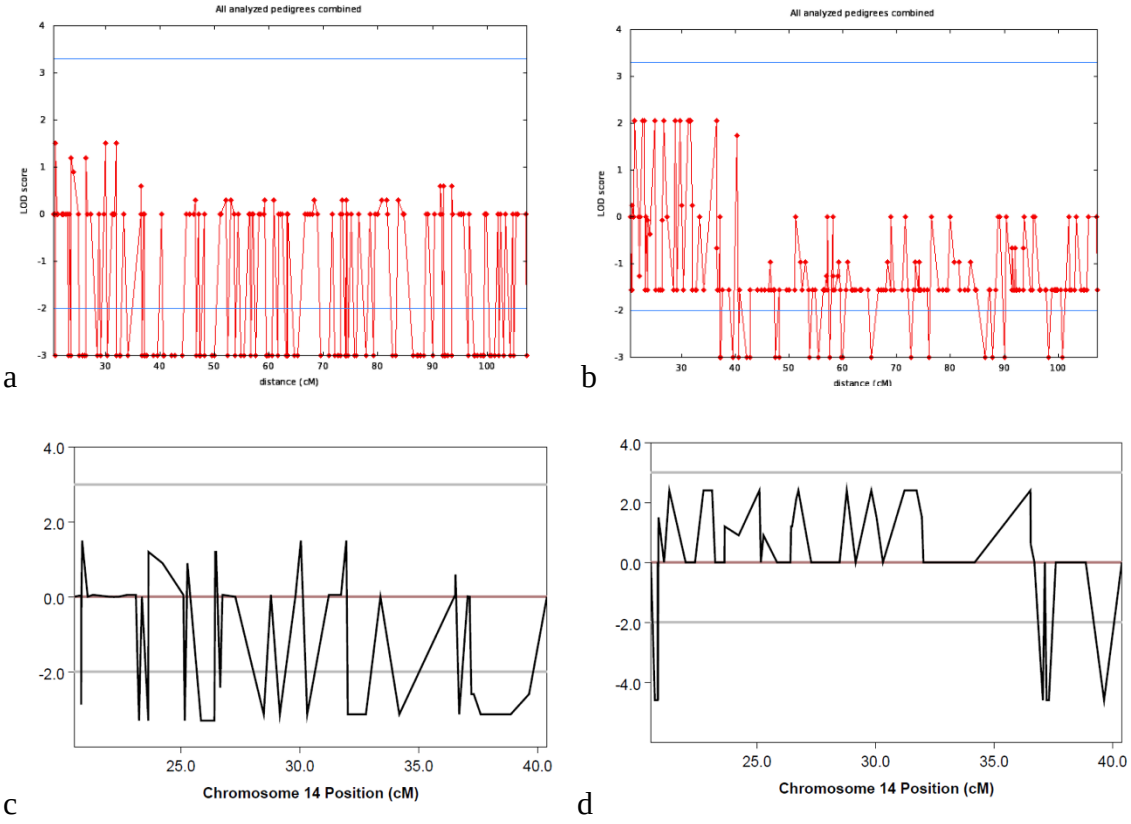


Figure 6. Linkage loci in family 6.
Parameters: omnibus affected, full penetrance, disease frequency: 0.001. Superlink graphs for dominant (a) and recessive mode of inheritance (b). MERLIN parametric linkage graphs for chromosome 14q11.2-13.3 (20.5-37.1 Mb) under dominant inheritance (c) and chromosome 14q11.2-21.1 (20.5-40.3 Mb) under recessive mode inheritance (d).

Table 3.
Highest single point LOD scores within each locus (SuperLink output).

Family	Locus	Inheritance Model	LOD score
1	4q31.23-q35.2	Recessive	2.4081
	4q31.23-q35.2	Dominant	1.1976
2	3p22.1-p14.2	Dominant	2.0249
	9q21.33-q33.1	Dominant	2.4519
	15q21.1-q25.2	Dominant	2.0267
	3p22.1-p14.2	Recessive	1.6996
	9q21.33-q33.1	Recessive	2.5087
	15q21.1-q25.2	Recessive	3.0568
4	6q21-q25.1	Dominant	1.2041
	10q21.1-q25.1	Dominant	1.5008
	12p12.3-q15	Dominant	1.4965
	6q21-q22.3	Recessive	2.6522
	10q21.1-q25.1	Recessive	2.6524
	12p13.31-p11.2	Recessive	2.6522
6	14q11.2-q21.1	Recessive	2.0559
	14q11.2-q13.3	Dominant	1.5008

Microsatellites Markers

The LOD scores for the microsatellite markers analyzed were extracted from the LOD score tables in Appendix A, and are accompanied by their physical positions and heterozygosity scores (Table 4).

Genotyping problems. Six microsatellite markers had genotyping problems. Genotyping problems were observed in four markers were seen in family 2 (D3S3560, D9S283, D9S1689 and D15S972) and in two markers in family 4 (D10S1667 and D12S799). These markers were dropped from the targeted linkage analysis performed in MERLIN. These problems could have could have occurred due to a number of reasons that will be investigated in the future.

Table 4.
Marshfield microsatellite marker information and single point LOD scores from MERLIN.

				LOD scores	
Family	Marker	Physical Position (bps)	Heterozygosity Score	Recessive model	Dominant model
1	D4S393	162521286-162521391	0.71	2.4081	1.1976
2	D3S3560	48194274-48194557	0.66	NA	NA
	D3S1588	54097848-54098209	0.75	-1.6882	-0.2057
	D9S283	92464311-92464628	0.81	NA	NA
	D9S1689	96504943-96505247	0.66	NA	NA
	D9S1683	113863195-113863536	0.76	-6.6712	-1.1674
	D9S289	116415496-116415772	0.75	1.1899	0.8722
	D15S1027	79118044-79118325	0.66	-4.9479	0.7914
	D15S972	85911589-85911882	0.78	NA	NA
4	D6S454	115377348-115377707	0.69	1.7975	-3.5831
	D6S2259	118935955-118936213	0.73	1.7974	0
	D10S1670	68874225-68874593	0.76	1.5008	0
	D10S1667	80864616-80864867	0.72	NA	NA
	D10S1679	123133495-123133851	0.57	0	-0.5035
	D12S1715	16696869-16697246	0.63	0	0.0001
	D12S799	22473638-22474120	0.65	NA	NA
6	D14S264	25279919-25280182	0.70	0.8988	0.895
	D14S1280	26655785-26656114	0.70	2.0956	-2.4248

NA = Not available (markers with genotyping problems)

Targeted Loci Multipoint and Nonparametric Linkage Analysis in MERLIN

Multipoint (3-point) linkage analysis was performed in MERLIN on the targeted loci and markers in all the reported loci showed higher or similar LOD scores (Table 5) compared to the single point LOD scores (Table 3). For multipoint linkage analysis, a LOD score above 2.0 was obtained under the recessive model for the microsatellite marker D4S393 on chromosome 4q in

family 1 and the marker D14S1280 on chromosome 14q in family 6 (Table 5). The highest LOD score obtained under dominant mode of inheritance was 1.19 at D4S393 and this marker was located next to the SNP marker that showed the highest LOD score on chromosome 4 in family 1 (Table 5). Multipoint LOD score could not calculated for 3p22.1-p14.2, 9q21.33-q33.1 and 15q21.1-q25.2 in family 2 due to the family size.

Table 5.
Highest multipoint LOD scores within each locus.

Family	Locus	Inheritance Model	LOD score
1	4q31.23-q35.2	Recessive	2.4081
	4q31.23-q35.2	Dominant	1.1976
2	3p22.1-p14.2	Dominant	NC
	9q21.33-q33.1	Dominant	NC
	15q21.1-q25.2	Dominant	NC
	3p22.1-p14.2	Recessive	NC
	9q21.33-q33.1	Recessive	NC
	15q21.1-q25.2	Recessive	NC
4	6q21-q25.1	Dominant	-1.7992
	10q21.1-q25.1	Dominant	-0.1584
	12p12.3-q15	Dominant	-1.5537
	6q21-q22.3	Recessive	1.7972
	10q21.1-q25.1	Recessive	1.797
	12p13.31-p11.2	Recessive	1.797
6	14q11.2-q21.1	Recessive	2.339
	14q11.2-q13.3	Dominant	-0.6038

NC = Not calculated

Nonparametric analysis. No significant or suggestive LOD scores were found under the nonparametric linkage analysis (Table A1 – A8). LOD scores obtained through nonparametric linkage analysis were consistently below 1.0, on all loci, which is lower than the threshold of significance (LOD score of 4.20) (Nyholt, 2000).

Discussion

Our linkage results in six SLI families indicate the power of the pedigree-based approach for complex phenotypes and provide supportive evidence for the variability of the behavioral

phenotype and genetic heterogeneity of SLI. Three linkage loci were identified in SLI family 2, another three chromosomal loci showed a linkage in family 4. Two SLI families, family 1 and 6 showed suggestive linkage.

SNP markers with the highest LOD scores were found under the recessive models of inheritance for family 1 and 4. This is not expected for outbred families. In outbred families, we expect that there has been numerous cross over or recombination events preventing the transmission of rare recessive disorders (Read & Strachan, 2011). The likelihood is very small that both parents would be carriers for a susceptible allele in outbred families. The highest LOD scores obtained under the recessive model in outbred families might indicate the existence of variable mode of inheritance SLI, consistent with the complexity.

Loci within Context of Previous Literature

The linkage loci identified in family 2 (chromosome 3p14.2-p22.1 and 15q24.1-q25.2), and a linkage locus (chromosome 6q21-q25.1) identified in family 4 are all found on chromosomes previously investigated by Rice and colleagues (2009) however the loci defined and presented here do not coincide.

As mentioned earlier, participants from the Rice et al. (2009) study are part of a larger ongoing longitudinal study in the LAS lab. Specifically, we know individuals from family 2 were the part of that study. The locus on chromosome 15q24.1-q25.2 is the closest to the loci previously investigated, about 13 Mb away, on chromosome 15q14-q21 (Rice et al., 2009). The locus on chromosome 3p22.1-p14.2 is greater than 40 Mb from the previously investigated region 3p12-q13 (Rice et al., 2009). One of the more studied RD regions is on chromosome 6p22, that carries the candidate gene *KIAA0319* (Harold et al., 2006; Paracchini, 2011; Rice et al., 2009). However, the chromosomal region we identified in family 4 is located on the q arm.

There is a report of linkage on chromosome 6q11.2-q12 with the phenotype specified as phonological coding dyslexia (Petryshen et al., 2001), but the region is located 40 Mb away from the locus we identified on chromosome 6q21-q25.1. Rice and colleagues (2009) investigated loci based on genetic studies performed with individuals who have RD (Rice et al., 2009). Therefore, we did not necessarily expect the loci to overlap with these regions. Nevertheless, the suggestive linkage on the same chromosomes could indicate that genes on these chromosomes are involved in overall cognition and therefore can influence language ability. It also indicated that completely different genes could be implicated in these behavioral phenotypes, despite commonly being lumped together in investigation and discussion. Another idea that emerges here is the possibility that these regions could be linked to causal pathways, which could explain the relationship between early language difficulties and later reading disorders.

The pericentromeric region on chromosome 3 has been linked with speech-sound disorder (SSD) and RD (Nopola-Hemmi et al., 2001; Stein et al., 2004). In these studies, the most significant LOD scores were found for these phenotypes: phonological memory, single-word decoding, and nonsense word reading. However, in the previous studies (of those discussed here) no LOD scores above 1.0 were found for any of the reading phenotypes (i.e. WRMT or GORT) on the chromosome 3 region, only for the categorical PPVT phenotype (Rice et al., 2009).

The previously investigated region on chromosome 15q14-21 has been most strongly associated with SSD (Stein et al., 2006). The speech and language deficits seen in individuals with SSD are consistent with many other neurodevelopmental disorders (Stein et al., 2006). Previously, the region on chromosome 15q was associated with 8 phenotypes, ranging from the CTOPP NWR task to the GFTA. LOD scores greater than 0.6 were found for both the

categorical and quantitative omnibus standard score and WRMT phenotypes (Rice et al., 2009). Previous results found on chromosome 15 and our findings on the same chromosome although in a different location may indicate that different genes but on the same chromosome are involved in speech and language related phenotypes. The region on chromosome 15q was previously reported in developmental stuttering (Raza et al., 2013).

Interestingly, there is a report of a balanced translocation of chromosome 10q24.1 with chromosome 15q21.1 in a male with severe language delay, with almost no receptive or expressive language (Ercan-Sencicek et al., 2012). The male with this translocation had CELF scores in the 1st and 2nd percentiles across subtests, despite average nonverbal intelligence (PIQ = 100 on the WIPPSI) when assessed at 3;0 and 4;1 years of age (Ercan-Sencicek et al., 2012). Chromosome 10q24.1, which was associated previously with severe language delay was located within the locus (chromosome 10q21.1-q25.1) identified in our SLI family 4. A deletion within the same locus associated with a novel genetic disorder causing both cognitive and behavioral deficits was found in a study with three families (Balciuniene et al., 2007). Specifically, there was a deletion of the chromosome 10q22.3-23.3 regions (Balciuniene et al., 2007). This region contains several low-copy repeats (LCRs), which affect chromosome stability (Balciuniene et al., 2007). The other four loci we identified were also reported in previous studies of individuals with SSD, auditory processing disorder, other cognitive delays and RD.

Another locus on chromosome 12 (12p11.2-p13.31) identified in family 4 under the recessive model has been linked to auditory processing deficit and SLI (Addis et al., 2010). The locus on chromosome 12p13.31-q14.3 was identified in a three-generation German pedigree (Addis et al., 2010). The individuals in this family were phenotyped with the NWR task and showed a maximum LOD score of 2.1 on chromosome 12 when all members were included in

analysis (Addis et al., 2010). Six candidate genes were sequenced from this region, and three rare nonsynonymous variants were found. One variant was found in three individuals (*CNTN1* on exon 12) which could be a possible future sequencing opportunity for our SLI family 4 (Addis et al., 2010). This report is especially relevant to the current study, though it found linkage with the NWR task, which was not the categorical phenotype used in the analyses within investigation. However, the participants in this study have completed the NWR task as a part of the CTOPP. Therefore, if this region is investigated further, categorical affectedness status could be defined based on the CTOPP. Interestingly, a method similar to that of the current study (using a single family) identified this locus.

A study of 183 families, utilizing sibling pairs, evaluated participants with the question of reading ability in children identified with Attention-Deficit/Hyperactivity Disorder (ADHD) and their shared genetics (Loo et al., 2004). Participants were assessed with the Peabody Individual Achievement Test-Revised (PIAT-R) and QTL linkage was performed with the individual subtests as phenotypes. This study reported the highest multipoint maximum LOD scores (MLS) above 1.0 found across the 404 genome-wide highly polymorphic markers, for the PIAT-R subtests. Two markers near our loci were found to have the highest MLS for the PIAT-R spelling subtest. Specifically, the microsatellite marker D9S1677, at chromosome 9q31.3, with a LOD score of 1.2 is located on the boundary of the locus 9q21.33-q33.1 we reported in family 2 (Loo et al., 2004). Also, the marker D4S426, on chromosome 4q35.2 (Loo et al., 2004), with a LOD score of 1.53 is physically located at the boundary of the locus we identified on chromosome 4q31.2-q35.2 in family 1.

Loo et al. (2004) investigated reading in children with ADHD because of the comorbidity of RD and ADHD diagnoses, which is estimated between 25-40% (Loo et al., 2004). There are

similar estimates of comorbidity of SLI and ADHD diagnoses (35-50%) (Redmond, 2004). The early hypotheses for the overlaps in diagnoses are underlying developmental delays. However, more recent reports indicate that children with ADHD are picked up for speech and language services more often than those without ADHD (Redmond, 2016). Loo's (2004) results are still relevant to the current study. Language phenotypes are often associated with other cognitive phenotypes (Rice & Tager-Flusberg, 2016), which can help narrow down regions specific to language, or identify regions associated with general cognition. Similarly, the locus found in family 6 on chromosome 14q11.2-21.1 under the recessive model overlaps with the previously identified microdeletions in children with cognitive delays that could be relevant to narrowing chromosomal loci specific to language impairment (Zampini, D'Odorico, Zanchi, Zollino, & Neri, 2012).

Power and Limitations

Though 60 participants could be called a limitation, compared to previous GWAS studies where large cohorts were used, the power of our study comes from the relatedness of the individuals and their extensive assessment, as well as the use of a parametric linkage analysis. The purpose of this study was in part to show the power of using pedigrees in gene mapping of complex polygenic and complex Mendelian disorders like SLI. Genetic heterogeneity is consistently reported across SLI populations, even among related individuals. In using pedigrees of individuals with well-defined SLI phenotypes, we hoped to reduce the influence of genetic heterogeneity in the search for rare genetic variations of large effect. In contrast, the power of GWAS studies comes from the large population cohorts (thousands of participants per study). GWAS studies are very effective in finding the associated risk alleles for common traits, for which it is comparatively less challenging to ascertain a large population. There is increased

power in using larger families, which is supported by this study. In the largest family, family 2, the loci we identified were found on the same chromosomes investigated in the previous investigation by Rice and colleagues (2009). Many of the individuals from family 2 were included in the previous analysis, further showing the power of that family, even as a part of a larger cohort.

There were no linkage loci were reported in two of the smallest families (3 and 5), which each have seven members. In the case of families 3 and 5, the small sample size and shared phenotypes among affected members both play a role in the lack of suggestive linkage loci. The proband in family 3 is affected according to eight of the phenotypic measures, but other family members are only affected according to one or two measures without as much overlap in measures across the family members. The lack of consistency in affectedness of multiple measures in family 5 could be preventing genetic analysis of this family from benefiting from a pedigree-focused approach. This is not true of family 5, in which the four affected full siblings share affectedness status on five of the measures, indicating the size of this pedigree is the more limiting factor for this family.

Future Directions

I plan to follow up this study with sequencing of rare protein-coding variations, found within the linkage loci reported here, filtered down from the whole exome sequencing data of the same families. Rare variations could then be used to find the burden in a population, by determining the rates in population-matched cases and controls. Smaller families, which are less effective for linkage analysis, could be a target of follow-up investigation of resulting rare variations or candidate genes, identified in large families through exome sequencing. Another possibility could be to select candidate genes, within the linkage regions reported here, based on

previous studies and the function and expression of candidates. For example, I might sequence candidate genes within the chromosome 12p13.31-q14.3 region, sequenced in the study mentioned of the German family with SLI, specifically *CNTN1* (chromosome 12q12) (Addis et al., 2010).

Conclusion

This is the first genome-wide linkage analysis of families coming from the longitudinal SLI study at KU. The results revealed multiple SLI loci in family 2 and 4 and one locus each in family 1 and 6. If multiple loci are linked to the same phenotype in the same family, it is likely that expression of multiple genes is contributing to the transmission of SLI in the same family. The results of this study suggest using comprehensively assessed individuals is most beneficial for the genetic investigation of the SLI phenotype, even when affected status is defined categorically for only one measure. We propose a polygenic and polyphenic effect in our SLI families. The lack of overlap in the linkage loci supports the genetic and phenotypic heterogeneity seen in previous genetic studies of SLI.

References

Abecasis, G. R., Cherny, S. S., Cookson, W. O., & Cardon, L. R. (2002). Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nature genetics, 30*(1), 97.

Abreu, P. C., Greenberg, D. A., & Hodge, S. E. (1999). Direct power comparisons between simple LOD scores and NPL scores for linkage analysis in complex diseases. *The American Journal of Human Genetics, 65*(3), 847-857.

Addis, L., Friederici, A. D., Kotz, S. A., Sabisch, B., Barry, J., Richter, N., . . . Pääbo, S. (2010). A locus for an auditory processing deficit and language impairment in an extended pedigree maps to 12p13. 31-q14. 3. *Genes, Brain and Behavior, 9*(6), 545-561.

Balciuniene, J., Feng, N., Iyadurai, K., Hirsch, B., Charnas, L., Bill, B. R., . . . Czapansky-Beilman, D. (2007). Recurrent 10q22-q23 deletions: a genomic disorder on 10q associated with cognitive and behavioral abnormalities. *The American Journal of Human Genetics, 80*(5), 938-947.

Bartlett, C. W., Flax, J. F., Logue, M. W., Smith, B. J., Vieland, V. J., Tallal, P., & Brzustowicz, L. M. (2004). Examination of potential overlap in autism and language loci on chromosomes 2, 7, and 13 in two independent samples ascertained for specific language impairment. *Human heredity, 57*(1), 10-20.

Bartlett, C. W., Flax, J. F., Logue, M. W., Vieland, V. J., Bassett, A. S., Tallal, P., & Brzustowicz, L. M. (2002). A major susceptibility locus for specific language impairment is located on 13q21. *The American Journal of Human Genetics, 71*(1), 45-55.

Bishop, D. V. (2006). What causes specific language impairment in children? *Current directions in psychological science, 15*(5), 217-221.

Bishop, D. V., & Hayiou-Thomas, M. E. (2008). Heritability of specific language impairment depends on diagnostic criteria. *Genes, Brain and Behavior, 7*(3), 365-372.

Bishop, D. V., North, T., & Donlan, C. (1996). Nonword repetition as a behavioural marker for inherited language impairment: Evidence from a twin study. *Journal of child Psychology and Psychiatry, 37*(4), 391-403.

Broman, K. W., Murray, J. C., Sheffield, V. C., White, R. L., & Weber, J. L. (1998). Comprehensive human genetic maps: individual and sex-specific variation in recombination. *The American Journal of Human Genetics, 63*(3), 861-869.

Clerget-Darpoux, F., & Elston, R. C. (2007). Are linkage analysis and the collection of family data dead? Prospects for family studies in the age of genome-wide association. *Human heredity, 64*(2), 91-96.

Cooper, D. N., Krawczak, M., Polychronakos, C., Tyler-Smith, C., & Kehrer-Sawatzki, H. (2013). Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Human genetics, 132*(10), 1077-1130.

Day, J. J., & Sweatt, J. D. (2011). Epigenetic mechanisms in cognition. *Neuron, 70*(5), 813-829.

Dunn, D. M., & Dunn, L. M. (2007). *Peabody picture vocabulary test: Manual*: Pearson.

Ercan-Sencicek, A. G., Wright, N. R. D., Sanders, S. J., Oakman, N., Valdes, L., Bakkaloglu, B., . . . Grigorenko, E. L. (2012). A balanced t (10; 15) translocation in a male patient with developmental language disorder. *European journal of medical genetics, 55*(2), 128-131.

Fisher, S. E., Vargha-Khadem, F., Watkins, K. E., Monaco, A. P., & Pembrey, M. E. (1998). Localisation of a gene implicated in a severe speech and language disorder. *Nature genetics, 18*(2), 168-170.

Fraser, A., Macdonald-Wallis, C., Tilling, K., Boyd, A., Golding, J., Davey Smith, G., . . . Ness, A. (2012). Cohort profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *International journal of epidemiology, 42*(1), 97-110.

Greenberg, D. A., Abreu, P., & Hodge, S. E. (1998). The power to detect linkage in complex disease by means of simple LOD-score analyses. *The American Journal of Human Genetics, 63*(3), 870-879.

Hanscombe, K. B., Trzaskowski, M., Haworth, C. M., Davis, O. S., Dale, P. S., & Plomin, R. (2012). Socioeconomic status (SES) and children's intelligence (IQ): In a UK-representative sample SES moderates the environmental, not genetic, effect on IQ. *PloS one, 7*(2), e30320.

Harold, D., Paracchini, S., Scerri, T., Dennis, M., Cope, N., Hill, G., . . . Owen, M. J. (2006). Further evidence that the KIAA0319 gene confers susceptibility to developmental dyslexia. *Molecular Psychiatry, 11*(12), 1085.

Jordan, B. (2014). Genes and non-mendelian diseases: dealing with complexity. *Perspectives in biology and medicine, 57*(1), 118-131.

Kang, C., & Drayna, D. (2011). Genetics of speech and language disorders. *Annual review of genomics and human genetics, 12*, 145-164.

Loo, S., Fisher, S. E., Francks, C., Ogdie, M., MacPhie, I. L., Yang, M., . . . Monaco, A. (2004). Genome-wide scan of reading ability in affected sibling pairs with attention-deficit/hyperactivity disorder: unique and shared genetic effects. *Molecular Psychiatry, 9*(5), 485.

Luciano, M., Evans, D., Hansell, N., Medland, S., Montgomery, G., Martin, N., . . . Bates, T. (2013). A genome-wide association study for reading and language abilities in two population cohorts. *Genes, Brain and Behavior, 12*(6), 645-652.

Neale, M., & Cardon, L. R. (2013). *Methodology for genetic studies of twins and families* (Vol. 67): Springer Science & Business Media.

Newbury, D. F., Winchester, L., Addis, L., Paracchini, S., Buckingham, L.-L., Clark, A., . . . Everitt, A. (2009). CMIP and ATP2C2 modulate phonological short-term memory in language impairment. *The American Journal of Human Genetics, 85*(2), 264-272.

NIDCD. (2015). Specific language impairment. Retrieved from https://www.nidcd.nih.gov/health/specific-language-impairment

Nopola-Hemmi, J., Myllyluoma, B., Haltia, T., Taipale, M., Ollikainen, V., Ahonen, T., . . . Widén, E. (2001). A dominant gene for developmental dyslexia on chromosome 3. *Journal of medical genetics, 38*(10), 658-664.

Nudel, R., Simpson, N. H., Baird, G., O'Hare, A., Conti-Ramsden, G., Bolton, P. F., . . . Francks, C. (2014). Genome-wide association analyses of child genotype effects and parent-of-origin effects in specific language impairment. *Genes, Brain and Behavior, 13*(4), 418-429.

Nyholt, D. R. (2000). All LODs are not created equal. *American journal of human genetics, 67*(2), 282.

Oetting, J. B., & Rice, M. L. (1993). Plural acquisition in children with specific language impairment. *Journal of speech, language, and hearing research, 36*(6), 1236-1248.

Ott, J., Kamatani, Y., & Lathrop, M. (2011). Family-based designs for genome-wide association studies. *Nature Reviews Genetics, 12*(7), 465.

Ott, J., Wang, J., & Leal, S. M. (2015). Genetic linkage analysis in the age of whole-genome sequencing. *Nature Reviews Genetics, 16*(5), 275.

Paracchini, S. (2011). Dissection of genetic associations with language-related traits in population-based cohorts. *Journal of neurodevelopmental disorders, 3*(4), 365.
Petryshen, T. L., Kaplan, B. J., Fu Liu, M., de French, N. S., Tobias, R., Hughes, M. L., & Leigh Field, L. (2001). Evidence for a susceptibility locus on chromosome 6q influencing phonological coding dyslexia. *American Journal of Medical Genetics Part A, 105*(6), 507-517.
Raza, M. H., Gertz, E. M., Mundorff, J., Lukong, J., Kuster, J., Schäffer, A. A., & Drayna, D. (2013). Linkage analysis of a large African family segregating stuttering suggests polygenic inheritance. *Human genetics, 132*(4), 385-396.
Raza, M. H., Mattera, R., Morell, R., Sainz, E., Rahn, R., Gutierrez, J., . . . Brewer, C. (2015). Association between rare variants in AP4E1, a component of intracellular trafficking, and persistent stuttering. *The American Journal of Human Genetics, 97*(5), 715-725.
Read, A. P., & Strachan, T. (2011). *Human Molecular Genetics* New York: Garland Science: Taylor & Francis Group.
Reader, R. H., Covill, L. E., Nudel, R., & Newbury, D. F. (2014). Genome-wide studies of specific language impairment. *Current behavioral neuroscience reports, 1*(4), 242-250.
Redmond, S. M. (2004). Conversational profiles of children with ADHD, SLI and typical development. *Clinical linguistics & phonetics, 18*(2), 107-125.
Redmond, S. M. (2016). Language impairment in the attention-deficit/hyperactivity disorder context. *Journal of speech, language, and hearing research, 59*(1), 133-142.
Rhead, B., Karolchik, D., Kuhn, R. M., Hinrichs, A. S., Zweig, A. S., Fujita, P. A., . . . Raney, B. J. (2009). The UCSC genome browser database: update 2010. *Nucleic acids research, 38*(suppl_1), D613-D619.
Rice, M. L. (2012). Toward epigenetic and gene regulation models of specific language impairment: looking for links among growth, genes, and impairments. *Journal of neurodevelopmental disorders, 4*(1), 27.
Rice, M. L. (2016). Specific language impairment, nonverbal IQ, attention-deficit/hyperactivity disorder, autism spectrum disorder, cochlear implants, bilingualism, and dialectal variants: Defining the boundaries, clarifying clinical conditions, and sorting out causes. *Journal of speech, language, and hearing research, 59*(1), 122-132.
Rice, M. L. (2017). Overlooked by public health: Specific language impairment. *Open Access Government*.
Rice, M. L. (2018). Specific language impairment (SLI) versus speech sound disorder (SSD). *Open Access Government*.
Rice, M. L., Haney, K. R., & Wexler, K. (1998). Family histories of children with SLI who show extended optional infinitives. *Journal of speech, language, and hearing research, 41*(2), 419-432.
Rice, M. L., & Hoffman, L. (2015). Predicting vocabulary growth in children with and without specific language impairment: A longitudinal study from 2; 6 to 21 years of age. *Journal of speech, language, and hearing research, 58*(2), 345-359.
Rice, M. L., Redmond, S. M., & Hoffman, L. (2006). Mean length of utterance in children with specific language impairment and in younger control children shows concurrent validity and stable and parallel growth trajectories. *Journal of speech, language, and hearing research, 49*(4), 793-808.

Rice, M. L., Smith, S. D., & Gayán, J. (2009). Convergent genetic linkage and associations to language, speech and reading measures in families of probands with specific language impairment. *Journal of neurodevelopmental disorders, 1*(4), 264.
Rice, M. L., & Tager-Flusberg, H. (2016). Language Phenotypes. In *Neurophenotypes* (pp. 227-243): Springer.
Rice, M. L., & Wexler, K. (1996). Toward tense as a clinical marker of specific language impairment in English-speaking children. *Journal of speech, language, and hearing research, 39*(6), 1239-1257.
Rice, M. L., & Wexler, K. (2001). Test of early grammatical impairment. *San Antonio, TX: The Psychological Corporation*.
Rice, M. L., Wexler, K., & Cleave, P. L. (1995). Specific language impairment as a period of extended optional infinitive. *Journal of speech, language, and hearing research, 38*(4), 850-863.
Rice, M. L., Zubrick, S. R., Taylor, C. L., Hoffman, L., & Gayán, J. (2018). Longitudinal Study of Language and Speech of Twins at 4 and 6 Years: Twinning Effects Decrease, Zygosity Effects Disappear, and Heritability Increases. *Journal of speech, language, and hearing research, 61*(1), 79-93.
Rutter, M. (2007). Gene–environment interdependence. *Developmental science, 10*(1), 12-18.
Scerri, T. S., Morris, A. P., Buckingham, L.-L., Newbury, D. F., Miller, L. L., Monaco, A. P., . . . Paracchini, S. (2011). DCDC2, KIAA0319 and CMIP are associated with reading-related traits. *Biological psychiatry, 70*(3), 237-245.
Silberstein, M., Weissbrod, O., Otten, L., Tzemach, A., Anisenia, A., Shtark, O., . . . Shalata, A. (2012). A system for exact and approximate genetic linkage analysis of SNP data in large pedigrees. *Bioinformatics, 29*(2), 197-205.
SLI Consortium. (2002). A genomewide scan identifies two novel loci involved in specific language impairment. *The American Journal of Human Genetics, 70*(2), 384-398.
SLI Consortium. (2004). Highly significant linkage to the SLI1 locus in an expanded sample of individuals affected by specific language impairment. *The American Journal of Human Genetics, 74*(6), 1225-1238.
Stein, C. M., Millard, C., Kluge, A., Miscimarra, L. E., Cartier, K. C., Freebairn, L. A., . . . Lewis, B. A. (2006). Speech sound disorder influenced by a locus in 15q14 region. *Behavior Genetics, 36*(6), 858-868.
Stein, C. M., Schick, J. H., Taylor, H. G., Shriberg, L. D., Millard, C., Kundtz-Kluge, A., . . . Freebairn, L. A. (2004). Pleiotropic effects of a chromosome 3 locus on speech-sound disorder and reading. *The American Journal of Human Genetics, 74*(2), 283-297.
Strachan, T., Goodship, J., & Chinnery, P. (2014). *Genetics and genomics in medicine*: Taylor & Francis.
Stromswold, K. (1998). Genetics of spoken language disorders. *Human biology*, 297-324.
Stromswold, K. (2000). The cognitive neuroscience of language acquisition. *The new cognitive neurosciences, 2*, 909-932.
Tallal, P., Hirsch, L. S., Realpe-Bonilla, T., Miller, S., Brzustowicz, L. M., Bartlett, C., & Flax, J. F. (2001). Familial aggregation in specific language impairment. *Journal of speech, language, and hearing research, 44*(5), 1172-1182.
Taylor, C. L., Christensen, D., Lawrence, D., Mitrou, F., & Zubrick, S. R. (2013). Risk factors for children's receptive vocabulary development from four to eight years in the Longitudinal Study of Australian Children. *PloS one, 8*(9), e73046.

Tomblin, J. B., Freese, P. R., & Records, N. L. (1992). Diagnosing specific language impairment in adults for the purpose of pedigree analysis. *Journal of speech, language, and hearing research, 35*(4), 832-843.

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of speech, language, and hearing research, 40*(6), 1245-1260.

van Heyningen, V., & Yeyati, P. L. (2004). Mechanisms of non-Mendelian inheritance in genetic disease. *Human molecular genetics, 13*(suppl_2), R225-R233.

Vernes, S. C., Newbury, D. F., Abrahams, B. S., Winchester, L., Nicod, J., Groszer, M., . . . Geschwind, D. H. (2008). A functional genetic link between distinct developmental language disorders. *New England Journal of Medicine, 359*(22), 2337-2345.

Villanueva, P., Newbury, D. F., Jara, L., De Barbieri, Z., Mirza, G., Palomino, H. M., . . . Palomino, H. (2011). Genome-wide analysis of genetic susceptibility to language impairment in an isolated Chilean population. *European Journal of Human Genetics, 19*(6), 687.

Villanueva, P., Nudel, R., Hoischen, A., Fernández, M. A., Simpson, N. H., Gilissen, C., . . . Francks, C. (2015). Exome sequencing in an admixed isolated population indicates NFXL1 variants confer a risk for specific language impairment. *PLoS genetics, 11*(3), e1004925.

Wexler, K. (1994). Optional infinitives, head movement and the economy of derivations. In N. Hornstein & D. Lightfoot (Eds.), *Verb movement* (pp. 305).

Whitehouse, A. J., Bishop, D. V., Ang, Q., Pennell, C. E., & Fisher, S. E. (2011). CNTNAP2 variants affect early language development in the general population. *Genes, Brain and Behavior, 10*(4), 451-456.

Wiszniewski, W., Hunter, J. V., Hanchard, N. A., Willer, J. R., Shaw, C., Tian, Q., . . . Patel, A. (2013). TM4SF20 ancestral deletion and susceptibility to a pediatric disorder of early language delay and cerebral white matter hyperintensities. *The American Journal of Human Genetics, 93*(2), 197-210.

Wright, M. J., & Martin, N. G. (2004). Brisbane adolescent twin study: outline of study methods and research projects. *Australian Journal of Psychology, 56*(2), 65-78.

Zampini, L., D'Odorico, L., Zanchi, P., Zollino, M., & Neri, G. (2012). Linguistic and psychomotor development in children with chromosome 14 deletions. *Clinical linguistics & phonetics, 26*(11-12), 962-974.

Appendix A
MERLIN LOD Score Output Tables

Table A 1.
LOD scores on chromosome 4q31.23-q35.2 in family 1

Physical Position (bps)	rsID	Single-point Recessive	Dominant	Non-parametric
150712923	rs1394845	1.2041	0.9233	0.567
151495331	rs991529	-2.8925	0.5969	0.565
152355268	rs2709828	0	0	0.565
152994226	rs17361055	-3.7955	-0.2942	0.565
153330301	rs1516822	1.2041	0.9233	0.565
154076697	rs4235229	-2.8925	0.5969	0.565
154887604	rs1878449	-3.7955	-0.2942	0.565
156506869	rs716428	0	0	0.565
157050382	rs4259017	1.2041	0.9233	0.566
157480074	rs2090870	1.2041	0.9233	0.566
157489906	rs1554472	0	0	0.566
158620303	rs11100128	1.2041	0.9233	0.566
160004853	rs510719	-2.8925	0.5969	0.565
161767814	rs6536541	2.4081	1.1976	0.897
161600134	D4S393	2.4081	1.1976	0.897
162733928	rs2014158	-3.7955	-0.2942	0.565
163390933	rs4139	0	0	0.565
163690348	rs1458149	-3.1935	0.2976	0.565
163781864	rs2054210	0	0	0.565
165075368	rs5020379	1.2041	0.9233	0.565
165281266	rs4318612	1.2041	0.9233	0.565
167079393	rs1039180	-3.1935	0.2976	0.565
167744429	rs954722	1.2041	0.9233	0.565
168150594	rs1385513	1.2041	0.9233	0.565
168437001	rs1478224	0	0	0.565
169663615	rs6811238	-3.7955	-0.2942	0.565
169732171	rs1566499	1.2041	0.9233	0.565
170939829	rs931145	0	0	0.565
171644884	rs1533837	-3.7955	-0.2942	0.565
171743216	rs2877558	-3.7955	-0.2942	0.565
172383439	rs2178299	-3.7955	-0.2942	0.565
173485263	rs1472370	1.2041	0.9233	0.565
173764577	rs1824347	1.2041	0.9233	0.565
173765727	rs13112941	1.2041	0.9233	0.565
174723835	rs4695886	-3.7955	-0.2942	0.565
174859997	rs3889926	1.2041	0.9233	0.574

Table A1 continued

Physical Position	rsID	Single-point		
		Recessive	Dominant	Non-parametric
175699276	rs2200457	0	0	0.623
176545639	rs13132745	-2.5915	0.897	0.667
176548782	rs1014381	-2.5915	0.897	0.667
177322096	rs12647397	0	0	0.667
178305908	rs12649669	-3.4945	0	0.667
178621097	rs1993281	1.2041	0.9233	0.667
178995242	rs1711393	1.2041	0.9233	0.667
179921499	rs2044868	0	0	0.667
179989221	rs12499500	-2.5915	0.897	0.667
180824638	rs1870347	-3.6194	-0.1226	0.667
181288587	rs10025005	-3.7955	-0.2942	0.667
181636440	rs1363207	-3.7955	-0.2942	0.667
182327943	rs724659	-3.7955	-0.2942	0.565
182944838	rs10011443	1.2041	0.9233	0.565
182979354	rs335077	1.2041	0.9233	0.565
183313353	rs2045405	1.2041	0.9233	0.565
183321916	rs1516538	1.2041	0.9233	0.565
183562256	rs3860640	-3.7955	-0.2942	0.565
183959226	rs2016910	0	0	0.565
184324023	rs907362	1.2041	0.9233	0.565
184600601	rs4241779	-3.7955	-0.2942	0.565
185119844	rs1921565	1.2041	0.9233	0.565
185955057	rs1564986	0	0	0.565
185981436	rs6844558	-3.7955	-0.2942	0.565
186316829	rs4862524	-3.1935	0.2976	0.565
186469493	rs1158465	-2.8925	0.5969	0.565
186732566	rs1566347	-3.7955	-0.2942	0.565
187099043	rs2036912	1.2041	0.9233	0.661
187122319	rs3736455	1.2041	0.9233	0.666
187127817	rs1473597	-3.4945	0	0.667
187421850	rs6553000	1.2041	0.9233	0.667

Table A 2.

LOD scores on chromosome 3p22.1-p14.2 in family 2.

Physical Position	rsID	Recessive single-point	Dominant single-point	Non- parametric
43518167	rs1351631	-7.7179	1.3922	0.065
43558085	rs737516	-7.7179	1.3922	0.063
43626375	rs1013758	-7.7179	1.3922	0.059
44105159	rs1008369	0.7856	-1.533	0.074
44753363	rs954282	-3.6271	-1.686	-0.003
45083133	rs2056321	-1.6882	-0.2057	-0.003
45276370	rs33759	-1.2593	-1.6547	-0.002
45923085	rs1860264	-4.9812	2.0995	-0.002
46352384	rs6441961	-5.4466	-2.0087	-0.002
46510209	rs1520483	-3.3556	-1.3937	-0.002
46556835	rs6808142	-2.1672	0.1219	-0.002
46556948	rs17030627	-6.055	-1.366	-0.002
46574418	rs1402152	-6.3211	-1.4979	-0.002
46902129	rs2227294	-1.6882	-0.2057	-0.002
47575568	rs4858833	0.3568	0.1645	-0.002
47652639	rs1014228	-3.8933	-2.3616	-0.002
47915118	rs319682	-3.8933	-2.3616	-0.002
48194274	D3S3560	0.3568	0.1645	-0.002
48541182	rs7434107	-6.4161	-3.602	-0.002
49362892	rs1865741	-4.9338	-2.4968	-0.001
50114515	rs7061	-4.7296	2.1175	-0.001
50201924	rs2188151	0.7856	0.1637	-0.001
50291785	rs2282751	-7.7159	-1.5825	-0.001
50371432	rs2236947	-3.6234	-1.9098	-0.002
51416368	rs11712780	-6.2925	-3.3244	-0.002
51581509	rs11720298	-3.2309	-1.3657	-0.002
52477866	rs11716487	-6.4161	-3.8086	-0.002
52727257	rs2289247	-6.055	-1.366	-0.002
52833805	rs3617	-2.1672	0.1219	-0.002
53498942	rs2101397	-5.3167	-1.6918	-0.002
53909731	rs893367	-3.2309	-1.3657	-0.002
54077256	rs9856661	-11.8459	-3.3348	-0.002
54097848	D3S1588	-1.6882	-0.2057	-0.002
54595678	rs9864433	0.3568	0.1645	-0.002
54628483	rs28594819	-7.8457	-1.7079	-0.002
54909196	rs4955903	-1.6882	-1.6537	-0.002
55271236	rs920891	-2.5193	1.9677	-0.002
56533016	rs978979	-6.841	-2.4316	-0.002
56796026	rs9825091	-1.6882	-1.6537	-0.002

Table A 3.

LOD scores on chromosome 9q21.33-q33.1 in family 2.

Physical Position	rsID	Recessive single-point	Dominant single-point	Non- parametric
87395810	rs3181360	0.9566	-0.4776	0.521
87861650	rs1167768	-3.5774	0.2884	0.524
88269947	rs729958	-5.1542	0.3451	0.504
89106020	rs912067	-4.1804	-0.9987	0.495
89961776	rs913444	0.232	-0.1291	0.485
90287187	rs884416	-0.8406	-0.2841	0.48
90583521	rs7046597	-4.5441	-0.2024	0.475
91154126	rs12338220	0	-0.0702	0.622
91899245	rs4534181	0	-0.0705	0.763
92298168	rs700962	-2.1672	-0.9491	0.824
92464311	rs1250019	-5.0683	-0.997	0.872
93080194	rs7870701	-3.1528	0.6778	0.883
93206737	rs4744136	-5.2359	0.3885	0.907
93567281	rs3001115	-4.3391	-0.4556	0.919
93794773	rs10820412	0	-0.0702	0.918
93941796	rs7025117	-4.5441	-0.2027	0.915
94369120	rs1335049	-0.9183	0.6378	0.904
94685926	rs1930243	-1.7422	-0.8874	0.902
94734562	rs675837	-2.1672	-2.7713	0.892
94974188	rs7042948	0.9581	0.2106	0.879
95275024	rs10821119	0	-0.0702	0.849
95849542	rs7022714	-5.0683	-0.8224	0.838
96093924	rs10978931	-1.566	-0.7063	0.837
96145271	rs17577532	-5.3608	-0.9098	0.822
96504943	rs1757096	-2.1672	-0.9491	0.763
96951145	rs10868791	0	-0.0702	0.761
97321127	rs1338121	-1.4412	-0.089	0.542
97333275	rs2000182	-4.5292	0.869	0.511
98340320	rs786990	-4.9399	-1.1354	0.511
98447883	rs1778970	-2.1672	-2.7713	0.511
98578048	rs1407850	-4.8705	-1.9504	0.511
98826369	rs952765	-6.8107	-1.5525	0.509
99330052	rs726657	-3.6519	-0.0573	0.907
100367306	rs88644	-1.5177	-1.3628	0.91
100607497	rs1819730	-3.0106	-0.0648	0.916
100911671	rs28412870	-1.7391	-0.7084	0.917
101331392	rs1000709	0.9566	0.7767	0.917
101565645	rs2780701	2.3825	0.4796	0.913
101855570	rs1250288	-2.669	-1.6207	0.913

Table A3 continued

Physical Position	rsID	Recessive single-point	Dominant single-point	Non- parametric
102845908	rs2001917	1.782	0.7848	0.908
102897144	rs3852401	0	-0.0702	0.898
103015018	rs4742762	0.232	-0.0702	0.896
103116172	rs1980889	0.6601	-0.7741	0.716
103140157	rs28453748	-5.3608	-2.7043	0.527
103880484	rs1932158	-1.7391	-2.8089	0.485
104133628	rs14419	-3.6519	0.0209	0.44
104417554	rs4598317	-0.8406	-0.2841	0.44
104686402	rs10081701	-2.669	-1.342	0.44
104792810	rs902488	0	-0.0702	0.462
105009468	rs1463983	-0.8406	-0.2841	0.509
105322593	rs949471	-3.5931	-1.2874	0.517
106058590	rs1017890	0.6601	-0.7741	0.52
106366275	rs6478437	-2.5334	-1.1077	0.513
107112483	rs4744070	-4.8396	-0.2028	0.513
107544700	rs1048510	-3.4595	0.2805	0.527
107562804	rs2246522	-0.3177	0.8705	0.943
107571241	rs10991231	-2.1672	-0.7744	0.978
108426784	rs4077800	0.232	-0.0705	0.98
108656238	rs2777804	-5.5332	-0.9422	0.988
108681783	rs1751798	-1.7391	-0.7082	0.991
108912911	rs1316268	-3.1528	0.2975	0.992
109073701	rs2622266	-1.4426	-0.686	0.992
109162816	rs363717	-5.2481	-0.9035	0.993
109354827	rs1323421	-1.7422	-0.8871	0.993
110231863	rs940287	-1.7391	-2.8089	0.993
110346728	rs1327532	-1.7391	-1.0002	0.987
110437175	rs2026999	0	-0.0705	0.985
110950572	rs10739288	-9.8435	-4.3619	0.963
111064801	rs3780528	0.6601	-0.6617	0.963
112078548	rs10759180	0.6601	-0.7744	0.963
112400265	rs1961970	-1.7391	-1.0002	0.961
112775268	rs984071	2.3841	1.4388	0.961
113126314	rs560019	-2.1672	-2.4175	0.968
113139030	rs1491100	-2.1672	-0.9491	0.964
113403846	D9S1683	-6.6712	-1.1674	0.951
113667277	rs10761323	-2.5334	-1.1077	0.5
113863195	rs1329088	-1.4412	-1.8648	0.859
114831896	rs1572983	-3.5931	-1.0172	0.965
115416257	rs2230808	-5.6762	0.094	0.982

Table A3 continued

Physical Position	rsID	Recessive single-point	Dominant single-point	Non- parametric
115971429	D9S289	1.1899	0.8722	0.993
116170559	rs1516882	-4.0212	-0.1813	0.993
116415496	rs3780346	-3.1528	0.6778	0.993
116420246	rs10819780	-2.1672	-0.9491	0.996
116501277	rs1555519	-1.4412	-0.089	0.996
117106522	rs1529191	-2.8867	0.3928	0.995
117237754	rs814027	0.232	-0.0705	0.995
117691558	rs28655161	0	-0.0702	0.987
117696336	rs1871692	-1.4426	-0.6862	0.987
117849641	rs1923433	-5.0683	-0.997	0.974

Table A 4.

LOD scores on chromosome 15q24.1-q25.2 in family 2.

Physical Position	rsID	Recessive single-point	Dominant single-point	Non- parametric
74105493	rs896588	-6.0562	-0.5806	0
74147244	rs1823718	-5.672	-0.3728	0
74581073	rs1484214	-1.595	-2.4477	0.381
74703929	rs741761	-3.7701	-0.1374	-0.002
74709566	rs2075590	-2.1672	0.2082	0.182
74709975	rs2075589	-2.1672	0.2082	0.182
74713300	rs11857558	0.3568	0.1059	0.146
74718699	rs11852760	-1.647	0.1456	0.138
74723644	rs1992145	-1.647	0.1456	0.138
74884447	rs12898794	-4.8666	0.0335	0.111
75012979	rs41279188	-4.0408	0.1871	-0.002
75041917	rs762551	0	0	0
75047426	rs2470890	0	0	0
76126371	rs12050778	0.6275	0.4681	0.212
76376116	rs2955736	-4.2112	-1.7777	0.668
76674624	rs744336	-1.7718	0.32	0.193
77335891	rs11636648	-1.7718	-1.6949	0
77540102	rs11639314	-3.3162	0.5728	-0.002
77926103	rs74025333	-4.0408	0.1871	-0.002
77984480	rs4243047	-2.1672	0.2082	0.182
78541769	rs1519819	0.6275	-0.2312	0
78789223	rs965604	0.9284	0.3309	0
79118044	D15S1027	-4.9479	0.7914	0.093
79292521	rs11072823	3.0568	2.0116	1.077
79439159	rs1402760	-6.1403	-0.0264	-0.002
80404656	rs4778752	-2.4674	-1.8343	0
80612431	rs2034247	0.6275	0.0829	0.212
80910857	rs11072930	-1.8119	-1.011	0.682
81112449	rs1553650	1.5743	-0.3235	0.005
81778698	rs4075641	1.0814	0.264	0.189
82021629	rs1567897	-6.865	-2.795	-0.002
82194085	rs12916134	-3.0455	-1.7173	-0.002
83215251	rs1267657	-5.3478	-0.9441	0.099
83507051	rs2046071	-8.1441	-1.6032	-0.002
83663941	rs2123157	-6.4401	-1.996	-0.002
83745534	rs11259964	-6.47	-1.9846	0.284
83959242	rs7180692	-9.2642	-1.9963	-0.002
84452052	rs1426165	0.232	0.2803	0.201
84573491	rs899926	-0.7881	0.6479	0.426

Table A 5.

LOD scores on chromosome 6q21-q22.3 in family 4.

Physical Position	rsID	Recessive single-point	Dominant single-point	Non-parametric
110635602	rs7748982	-7.2021	0	0.749
111260923	rs781499	0	-2.2015	0.758
111283592	rs9374227	-3.4064	-3.3996	0.758
111427286	rs1022092	-7.2021	0	0.758
111870090	rs4947122	0	-2.2015	0.758
112500650	rs1158747	0	-2.2015	0.758
113151042	rs4945917	-7.2021	0	0.758
113666000	rs3851197	-3.4065	-3.3996	0.758
113669875	rs1565528	-3.4065	-3.3996	0.758
113683910	rs1491074	0	-2.2015	0.758
113760063	rs1033391	-7.2021	0	0.758
114049192	rs773676	1.7974	0.0001	0.897
114067127	rs2030926	0	0	0.897
114691511	rs1415428	0	0	0.897
115377348	D6S454	1.7975	-3.5831	0.897
115779659	rs7759765	0	-2.2015	0.897
116076012	rs9374570	1.2041	1.2041	0.897
116520348	rs1204842	0	0.0001	0.822
116689378	rs1041883	0	0	0.79
117608318	rs544047	0.5934	0.5935	0.552
117684992	rs1321807	0	0	0.552
117724462	rs2243379	0	0	0.552
117860058	rs210617	1.7974	0.0001	0.552
118023316	rs1541317	1.7975	0.0001	0.552
118935955	D6S2259	1.7974	0	0.552
119460786	rs9372523	0	0	0.552
119508207	rs1012509	0	-4.7047	0.552
119833118	rs7758258	1.5008	0	0.552
120194916	rs937091	1.7974	0.0001	0.552
121398535	rs218867	1.7975	0.0001	0.551
122479746	rs1563512	0	0	0.55
123724482	rs873460	1.7974	0	0.549

Table A 6.

LOD scores on chromosome 10q21.1-q25.1 in family 4.

Physical Position	rsID	Recessive single-point	Dominant single-point	Non-parametric
55683138	rs1900423	1.7976	0.0001	0.848
56001078	rs996320	1.7976	0.0001	0.852
56847522	rs1733743	1.7975	0.0001	0.861
57629679	rs1338799	0	0.0001	0.869
57650954	rs10825659	0	0.0001	0.87
59823894	rs2184033	1.7975	0.0001	0.895
60012231	rs2590339	0	-2.2015	0.897
60262925	rs1427209	0	0.0001	0.897
60348886	rs10740731	0.2925	0.2926	0.897
60571435	rs4948317	1.7975	0.0001	0.897
60896265	rs1897620	1.7976	0.0001	0.897
61081713	rs284642	0.0001	0	0.897
61085082	rs1999668	0.0001	0	0.897
61790383	rs12357206	0	0	0.897
61791039	rs7911953	0	0	0.897
62661961	rs2893869	0	-0.2024	0.897
62933123	rs1906455	0	-0.2025	0.897
62938418	rs1906457	0	-0.2025	0.897
63814914	rs10821951	0	-0.2024	0.897
64470675	rs224136	1.7976	0.0001	0.897
64516065	rs377859	0	0	0.897
64620042	rs911610	1.7976	0.0001	0.897
64936679	rs907	0	0.0001	0.897
65253700	rs7910662	0	0.0001	0.897
65823789	rs1866311	0	-0.2024	0.897
67282121	rs7099767	0	-0.2025	0.897
68874225	D10S1670	1.5008	0.0001	0.897
68961492	rs1904645	0	-0.2024	0.897
69023117	rs1904610	1.5009	1.5008	0.897
69664959	rs10997868	1.7976	0.0001	0.897
71158268	rs1227938	1.7976	0.0001	0.897
71226790	rs736594	1.7976	0.0001	0.897
71868763	rs2271698	0	0	0.897
72744630	rs756322	0	-0.2025	0.897
72813306	rs10762420	0	0	0.897
72824456	rs10999657	1.7976	0.0001	0.897
73315940	rs877783	1.7976	0.0001	0.897
73335042	rs1867998	1.7975	0.0001	0.897
73769543	rs4148944	0	-0.2024	0.897

Table A6 continued

Physical Position	rsID	Recessive single-point	Dominant single-point	Non-parametric
73770073	rs4148946	1.7976	0.0001	0.897
73770651	rs4148949	1.7976	0.0001	0.897
73771706	rs4148950	0.0001	0	0.897
73772014	rs1871450	0.0001	0	0.897
73772336	rs731027	0.0001	0	0.897
73772762	rs730720	0.0001	0	0.897
74402802	rs3000976	1.7975	0.0001	0.897
75300994	rs4746136	0	0.0001	0.897
75654931	rs2688610	1.5009	1.5008	0.897
75669190	rs2227551	0	-0.2024	0.897
75676464	rs4065	1.5009	1.5008	0.897
76180335	rs11001034	0	0.0001	0.897
76787128	rs1551067	0	0.0001	0.896
77504287	rs7097617	1.7975	0.0001	0.896
77646553	rs9415126	1.7976	0.0001	0.896
78122729	rs12766217	0	0.0001	0.896
78299651	rs1907308	1.7976	0.0001	0.896
79182034	rs158421	1.4966	0.0001	0.896
80864867	D10S1667	NA	NA	0.896
80930439	rs703990	1.7976	0.0001	0.896
81103862	rs1892498	1.7975	0.0001	0.896
81164521	rs2279335	-7.2021	0.0001	0.757
82033470	rs7087728	0	0	0.897
82035560	rs4934027	1.7975	0.0001	0.897
82472563	rs720262	0	-0.2025	0.897
82568684	rs10882097	0	-0.2024	0.897
82832257	rs1336439	0	-0.2024	0.897
83723457	rs7902158	0	0	0.897
83811949	rs11192313	1.7976	0.0001	0.897
84074976	rs7069367	0	0	0.897
84639649	rs2475793	0	-0.2024	0.897
85221278	rs4600152	1.4966	0.0001	0.897
85456128	rs4933299	1.7975	0.0001	0.897
86323214	rs1188786	1.5008	1.5008	0.897
86633900	rs7910550	1.7974	0.0001	0.897
87516345	rs1880382	1.7975	0.0001	0.897
87847009	D10S1769	0	-0.5035	0.897
87862479	rs9420382	0.0001	-0.2027	0.897
88079367	rs10887577	0	-0.2025	0.897
88461699	rs10887650	0	-0.2024	0.897

Table A6 continued

Physical Position	rsID	Recessive single-point	Dominant single-point	Non-parametric
88716577	rs10887683	1.5008	1.5008	0.897
88730312	rs4869	1.5008	1.5008	0.897
89336834	rs7914810	0	-0.2024	0.897
89987183	rs2039305	0	-5.0619	0.552
90195149	rs1935581	1.4966	0.0001	0.634
91203214	rs1556611	0	-0.2024	0.897
92090566	rs1001065	1.7976	0.0001	0.897
92304910	rs7094359	0	-0.2025	0.897
92470777	rs1857586	0	0	0.897
92944658	rs3980838	0	0.0001	0.897
93399130	rs7475251	1.7976	0.0001	0.897
94462882	rs1111875	0	0.0001	0.897
94839724	rs4418728	1.5009	1.5008	0.897
94921065	rs4918664	0	0	0.897
94936328	rs4244304	0	0	0.897
95036575	rs787652	0.0001	0	0.897
95190568	rs701873	0	-0.2025	0.897
96495232	rs2860840	0	-0.2025	0.897
96563757	rs4494250	0	-0.2025	0.897
96798749	rs10509681	0	0	0.897
97137105	rs1536556	0.0001	0	0.897
97172595	rs1410059	1.7975	0.0001	0.897
98279053	rs9093	0	0.0001	0.833
98288066	rs713251	0	0.0001	0.833
98699136	rs4919060	0	0.0001	0.806
99395652	rs60731885	-7.2021	0.0001	0.758
99637578	rs531676	-7.2021	0	0.758
99715744	rs489062	0	0	0.758
100268520	rs1889974	0	-0.2025	0.758
101563815	rs2273697	0	0	0.758
101798346	rs2862928	-1.7034	-1.7032	0.758
101846173	rs1410079	-1.7033	-1.7032	0.758
102104521	rs735877	1.1998	1.1998	0.758
102621414	rs11190730	0	-0.2024	0.758
103315801	rs1045232	0	0.0001	0.758
103519784	rs749694	0	-0.2024	0.758
104596924	rs6163	1.1999	1.1998	0.758
104597152	rs743572	1.1999	1.1998	0.758

Table A 7.

LOD scores on chromosome 12p13.31-p11.2 in family 4.

Physical Position	rsID	Recessive single-point	Dominant single-point	Non-parametric
8090703	rs933552	-0.2025	0	0.866
8976780	rs11047443	0	0	0.876
9262727	rs226386	-0.2025	0	0.88
9273711	rs226376	-0.2025	0	0.88
9822952	rs1560011	-0.2026	0	0.886
10560957	rs2617170	0	0	0.895
10772983	rs1870194	0.0001	-2.2016	0.897
10999780	rs1063193	0.0001	-2.2016	0.897
11701488	rs2416791	0	0	0.897
11812330	rs916041	-0.2026	0	0.897
11967710	rs732868	0	0	0.897
12446933	rs4763797	-0.2026	0	0.897
13054317	rs1548837	0.0001	0.0001	0.897
14138035	rs1895056	0	0	0.897
14653867	rs2900333	-0.4991	-0.4992	0.897
14982352	rs3088190	-0.2024	0	0.897
15577635	rs993123	0	0	0.897
16062450	rs10846247	-0.2024	0	0.897
16543935	D12S1715	0	0	0.897
16563973	rs1861577	0	0	0.897
16600984	rs993694	0.2923	0.2925	0.897
16650161	rs992690	0	-2.2015	0.897
17028884	rs10846448	0	0	0.897
19028431	rs1388093	0.2924	0.2925	0.897
19622345	rs1875467	1.7975	0.0001	0.897
20154110	rs10743347	0	0	0.897
21011480	rs4149117	1.7974	0.0001	0.897
21015760	rs7311358	1.7974	0.0001	0.897
21054369	rs3764006	1.7974	0.0001	0.897
21329738	rs2306283	0	-2.2015	0.897
21331549	rs4149056	0	-2.5012	0.897
21331599	rs4149057	0	-2.2015	0.897
21331625	rs2291075	0	-2.2015	0.897
21457434	rs11568563	0	0	0.897
21669027	rs1366	1.2041	1.2041	0.897
22320704	D12S799	0	0	0.897
22325245	rs2009625	0	0	0.897
23840513	rs6487356	0	-2.2015	0.897
24501160	rs575608	1.7975	0.0001	0.897

Table A7 continued

Physical Position	rsID	Recessive single-point	Dominant single-point	Non-parametric
24749874	rs7298579	0.2924	0.2925	0.897
25310724	rs7308865	1.7974	0.0001	0.805
25432626	rs7303282	1.7975	0.0001	0.782
25768113	rs11048128	0.0001	-0.0002	0.711
25861867	rs1381802	0	0	0.689
26222745	rs1546550	0	0.0001	0.592
26340888	rs875295	0.903	0.9031	0.555
26438189	rs6487543	0.903	0.9031	0.555
26697411	rs10842753	0	0	0.555
27085287	rs1151048	0	0.0001	0.555
27811224	rs7973582	0	0.0001	0.556
27940158	rs4244857	1.7975	0.0001	0.556

Table A 8.

LOD scores on chromosome 14q11.2-q21.1 in family 6.

Physical Position	rsID	Recessive single-point	Dominant single-point	Non- parametric
20532379	rs944398	0	0	0.671
20694602	rs1959344	-4.6	0.0231	0.673
20815013	rs1713430	-4.6	0.0231	0.674
20818131	rs1760921	0	-2.8789	0.674
20825965	rs2700	-4.6	0.0231	0.674
20852029	rs1760904	1.5008	1.497	0.674
21086092	rs1756370	0	0	0.82
21307824	rs11628338	2.3996	0.0485	0.924
21994936	rs2242529	0	0	0.924
22187832	rs8013476	0	-0.003	0.924
22390103	rs4981390	0	0	0.865
22747464	rs4982599	2.3996	0.0485	0.762
22782377	rs11845134	2.3996	0.0485	0.753
23113384	rs3102229	2.3996	0.0485	0.67
23242828	rs1061040	0	-3.3041	0.643
23368185	rs1570342	0	0	0.643
23636757	rs2268877	0	-3.3041	0.643
23640794	rs2268873	1.1998	1.196	0.643
24234400	rs9323367	0.8988	0.8951	0.643
25107565	rs1957528	2.3996	0.0485	0.643
25167511	rs12879771	0	-3.3041	0.643
25850832	D14S264	0.8988	0.895	0.643
25880588	rs1454361	0	-3.3041	0.643
26088391	rs941738	0	-3.3041	0.643
26093929	rs1956775	0	-3.3041	0.643
26412416	rs17461158	0	-3.3041	0.643
26446421	rs1950948	0	-3.3041	0.643
26486717	rs1956616	1.1998	1.196	0.643
26751099	rs862952	1.1998	1.196	0.643
27290857	D14S1280	2.0986	-2.4248	0.643
28488693	rs911085	2.3996	0.0485	0.648
28788618	rs1951085	0	0	0.676
29162832	rs880433	0	-3.1353	0.733
29824406	rs8011192	2.3996	0.0485	0.733
30045978	rs176264	0	-3.1353	0.733
30312091	rs1957115	2.3996	0.0485	0.733
31224458	rs11984	1.5008	1.497	0.733
31381351	rs1958980	0	-3.1353	0.733
31506505	rs179562	2.3996	0.0485	0.733

Table A8 continued

Physical Position	rsID	Recessive single-point	Dominant single-point	Non - parametric
31733642	rs2273171	2.3996	0.0485	0.733
31958483	rs7160965	2.3996	0.0485	0.733
32021438	rs727675	2.3996	0.0485	0.733
32636261	rs6571428	1.5008	1.497	0.733
32780576	rs1952896	0	-3.1353	0.733
33390916	rs2183192	0	-3.1353	0.733
34183707	rs2891193	0	-3.1353	0.733
36534827	rs7141672	0	0	0.732
36547931	rs1958053	0	-3.1353	0.732
36700375	rs368181	2.3996	0.0485	0.733
37059525	rs376927	0.5977	0.5942	0.733
37138628	rs1755784	0	-3.1353	0.539

Appendix B
Microsatellite Marker Reaction Mixture and Protocol

Table B 1.
Reaction mixture for amplification of microsatellite markers for genotyping.

Ingredients	Stock	Required amount
dH_2O	NA	4.3 uL
10x PCR Buffer	10x	1.0 uL
$MgCl_2$	25 mM	0.4 uL
dNTPs	10 nM	0.2 uL
genomic DNA	15 ng/uL	2.0 uL
Primer (forward)	8 mM/uL	1.0 uL
Primer (reverse)	8 mM/uL	1.0 uL
hot start Taq polymerase	5 units/uL	0.1 uL

Table B 2.
Thermocycling program for amplification of microsatellite markers.

95°C	15 minutes	HOLD
94°C	20 seconds	35 cycles
56°C	30 seconds	
72°C	1 minute	
72°C	7 minutes	HOLD
20°C	∞	HOLD

Protocol

1. Make the LIZ cocktail. Combine 1 mL of HiDi Formamide with 5 uL of ABI GeneScan 500 LIZ Size Standard in a 1.5 uL Eppendorf tube. Mix by inverting.
2. Preparing the plate:
    1. Following the full thermocycling program dilute the 10uL reactions with 25 uL of dH_2O (diluted plate)
    2. Then, transfer 2.5 uL from the diluted plate to a new plate containing 55 uL of dH_2O (pooled plate)
    3. Finally, make the loading plate. Put 9 uL of LIZ cocktail in each reaction well and combine with 1.3 uL of the pooled plate diluted reaction.